A Confidentiality Model for Ontologies

P. A. Bonatti and Luigi Sauro

Dept. of Electrical Engineering and Information Technologies
Università di Napoli "Federico II"

Abstract. We illustrate several novel attacks to the confidentiality of knowledge bases (KB). Then we introduce a new confidentiality model, sensitive enough to detect those attacks, and a method for constructing secure KB views. We identify safe approximations of the background knowledge exploited in the attacks; they can be used to reduce the complexity of constructing secure KB views.

1 Introduction

There is ample evidence of the need for knowledge confidentiality measures. OWL and the LOD paradigm are increasingly being used to encode the private knowledge of companies and public organizations. Linked open government data include potentially sensitive information, e.g. related to health. Medical records are annotated with semantic metadata based on SNOMED. FOAF assertions and other semantic description of social networks may affect the privacy of individuals. In all of these cases, semantic web techniques help in linking different knowledge sources and extract implicit information, thereby increasing security and privacy risks. Even the authors of *public* ontologies may want to hide some axioms to capitalize on their formalization efforts. See [8] for further motivations. In order to tackle the confidentiality requirements arising from these scenarios, several approaches have been proposed. The most popular security criterion is that the published view of the knowledge base should not entail any secret sentence (we call it *simple confidentiality model*). However, there exist attacks that cannot be blocked this way. The user may exploit various sources of background knowledge and metaknowledge to reconstruct the hidden part of the knowledge base. This paper contributes to the area of knowledge base confidentiality in several ways:

(i) It highlights some vulnerabilities of the approaches that can be found in the literature, including attacks based on meta-reasoning (Sec. 3).

(ii) It introduces a stronger confidentiality model that takes both object-level and meta-level background knowledge into account (Sec. 4), and it defines a method for computing secure knowledge views (Sec. 5) that generalizes some previous approaches.

(iii) It proposes a safe approximation of background metaknowledge (Sec. 6 and 7).

(iv) It investigates the computational complexity of constructing secure knowledge base views with our methodology (Sec. 7).

The paper is closed by a discussion of related work (Sec. 9), and conclusions. Some proofs are omitted due to space limitations.

2 Preliminaries on Description Logics

We assume the reader to be familiar with description logics, and refer to [1] for all definitions and results. We assume a fixed, denumerable signature Σ specifying the names

of *concepts*, *roles*, and *individuals*. Our framework is compatible with any description logic DL that enjoys compactness (needed by Theorem 6) and has decidable reasoning problems (e.g., $\mathcal{ALC}$, $\mathcal{EL}$, $\mathcal{SHIQ}$, etc.). We simply assume that our reference logical language $\mathcal{L}$ is generated from Σ by the grammar of the selected logic DL. By *axioms*, we mean members of $\mathcal{L}$, unless stated otherwise. A *knowledge base* is any subset of $\mathcal{L}$.[1]

Recall that axioms are expressions of the form $C \sqsubseteq D$, $R \sqsubseteq S$, $C(a)$, and $R(a, b)$ where C, D are concept expressions, R, S are role expressions, and a, b are individual constants. In some DL, an individual constant a may occur also in a *nominal*, that is, a concept expression $\{a\}$ denoting the singleton containing a. The axioms involving $\sqsubseteq$ are called *inclusions* (or *subsumptions*), while $C(a)$ and $R(a, b)$ are called *assertions*. In the simplest case, C and R are first order predicates and assertions are actually standard first-order atomic formulae. Inclusions are syntactic variants of logical implications.

The notion of *logical consequence* is the classical one; for all $K \subseteq \mathcal{L}$, the logical consequences of K will be denoted by $Cn(K)$ ($K \subseteq Cn(K) \subseteq \mathcal{L}$).

3 A simple confidentiality model

The most natural way of preserving confidentiality in a knowledge base KB is checking that its answers to user queries do not entail any secret. Conceptually, the queries of a user u are answered using u's view KB_u of the knowledge base, where KB_u is a maximal subset of KB that entails no secret. In order to illustrate some possible attacks to this mechanism, let us formalize the above *simple confidentiality model* (SCM).[2] It consists of: the knowledge base KB ($KB \subseteq \mathcal{L}$); a set of users U; a *view* $KB_u \subseteq KB$ for each $u \in U$; a set of *secrecies* $S_u \subseteq \mathcal{L}$ for each $u \in U$. Secrecies are axioms that may or may not be entailed by KB; if they do, then they are called *secrets* and must not be disclosed to u. Revealing that a secrecy is *not* entailed by KB is harmless [4]. For example, there is no need to protect the information that someone is *not* having chemotherapy.

A view KB_u is *secure* iff $Cn(KB_u) \cap S_u = \emptyset$. A view KB_u is *maximal secure* if it is secure and there exists no K such that $KB_u \subset K \subseteq KB$ and $Cn(K) \cap S_u = \emptyset$.

Attacks using object-level background knowledge. Frequently, part of the domain knowledge is not axiomatized in KB, therefore checking that $Cn(KB_u) \cap S_u = \emptyset$ does not suffice in practice to protect confidentiality. For example, suppose that there is one secret $S_u = \{OncologyPatient(John)\}$ and $KB_u = \{SSN(John, 12345), SSN(user123, 12345), OncologyPatient(user123)\}$. KB_u does not entail $OncologyPatient(John)$, so according to the SCM model KB_u is secure. However, it is common knowledge that a SSN uniquely identifies a person, then the user can infer that $John = user123$, and hence the secret.

In other examples, the additional knowledge used to infer secrets may be stored in a public ontology or RDF repository, and confidentiality violations may be automated.

Attacks to complete knowledge. Suppose the attacker knows that KB has complete knowledge about a certain set of axioms. Then the attacker may be able to reconstruct some secrets from the "I don't know" answers of a maximal secure view KB_u.

[1] Real knowledge bases are finite, but this restriction is not technically needed until Sec. 7.

[2] This usage of term "*model*" is common in Security & Privacy.

Example 1. Consider an organization's knowledge base that defines a concept *Employee* and a role *works_for* that describes which employees belong to which of the n departments of the company, $d_1, \ldots, d_n$. The *KB* consists of assertions like:

$$Employee(e) \qquad (1) \qquad\qquad works\_for(e, d_i) \qquad (2)$$

where we assume that each employee e belongs to exactly one department d_i. A user u is authorized to see all assertions but the instances of (2) with $i = n$, because d_n is a special department, devoted to classified projects. So S_u (the set of secrecies for u) is the set of all assertions $works\_for(e, d_n)$.

Note that there is one maximal secure view KB_u. It consists of all instances of (1), plus all instances of (2) such that $i \neq n$. Clearly, KB_u is secure according to SCM (because $Cn(KB_u) \cap S_u = \emptyset$). However, observe that $works\_for(e, d_n) \in Cn(KB)$ iff $Employee(e) \in Cn(KB_u)$ and for all $i = 1, \ldots, n$, $works\_for(e, d_i) \notin Cn(KB_u)$ (that is, the members of d_n are all the employees that apparently work for no department). Using this property (based on the knowledge that for each employee e, *KB* contains exactly one assertion $works\_for(e, d_i)$) and the knowledge of the protection mechanism (i.e. maximal secure views), that we assume to be known by attackers by *Kerchoff's principle*, a smart user can easily identify all the members of d_n. □

In practice, it is not hard to identify complete knowledge. A hospital's *KB* is expected to have complete knowledge about which patients are in which ward; a company's *KB* is likely to encode complete information about its employees, etc.

Some approaches filter query answers rather than publishing a subset of *KB* [7, 14, 16]. We call our abstraction of this method *simple answer confidentiality model* (SACM). It is obtained from the SCM by replacing the views $KB_u \subseteq KB$ with *answer views* $KB^a_u \subseteq Cn(KB)$. The difference is that KB^a_u is not required to be a subset of *KB* and—conceptually—KB^a_u may be infinite. KB^a_u is *secure* iff $Cn(KB^a_u) \cap S_u = \emptyset$.

The reader may easily verify that the SACM is vulnerable to the two kinds of attacks illustrated for the SCM. It is also vulnerable to a third kind of attacks, illustrated below.

Attacks to the signature. Suppose the user knows the signature of *KB* well enough to identify a symbol σ that does not occur in *KB*. First assume that σ is a concept name. It can be proved that:

Proposition 1. *If KB^a_u is a maximal secure answer view and σ is a concept name not occurring in KB, then for all secrecies $C \sqsubseteq D \in S_u$, $KB^a_u \models C \sqcap \sigma \sqsubseteq D$ iff $KB \models C \sqsubseteq D$.*

The problem is that although $C \sqcap \sigma \sqsubseteq D$ does not entail the secret inclusion $C \sqsubseteq D$, still a smart user knows that the former inclusion cannot be proved unless *KB* entails also the latter (then maximal secure answer views generally fail to protect secrets). This attack can be easily adapted to the case where σ is a role name. In practice, it is not necessary to be sure that σ does not occur in *KB*. The attacker may make a sequence of educated guesses (say, by trying meaningless long strings, or any word that is clearly unrelated to the domain of the *KB*); after a sufficient number of trials, the majority of answers should agree with the "real" answer with high probability. Rejecting queries whose signature is not contained in *KB*'s signature mitigates this kind of attacks but it leaks *KB*'s signature and it does not provide a complete solution: Any σ occurring in *KB* that is logically unrelated to C and D can be used for a similar attack.

4 A meta-safe confidentiality model

In this section we introduce a confidentiality model that makes the vulnerabilities illustrated above visible, by taking into account object- and meta-level background knowledge. A *bk-model* $\mathcal{M} = \langle KB, U, f, \langle S_u, PKB_u, BK_u \rangle_{u \in U} \rangle$ consists of a knowledge base $KB \subseteq \mathcal{L}$, a set of users U, plus:

- a *filtering function* $f : \wp(\mathcal{L}) \times U \to \wp(\mathcal{L})$, mapping each knowledge base K and each user u on a view $f(K, u) \subseteq Cn(K)$;
- for all $u \in U$:
  - a finite set of secrecies $S_u \subseteq \mathcal{L}$;
  - a set of axioms $BK_u \subseteq \mathcal{L}$, encoding the users' object-level knowledge;
  - a set of *possible knowledge bases* $PKB_u \subseteq \wp(\mathcal{L})$ (users' metaknowledge).[3]

The view of KB released to a user u is $f(KB, u)$. We adopt PKB because at this stage we do not want to tie our framework to any specific metalanguage. PKB represents the knowledge bases that are compatible with the user's metaknowledge.

Definition 1. *A filtering function f is* secure *(w.r.t. $\mathcal{M}$) iff for all $u \in U$ and all $s \in S_u$, there exists $K \in PKB_u$ such that:*

1. $f(K, u) = f(KB, u)$;
2. $s \notin Cn(K \cup BK_u)$.

Intuitively, if f is safe according to Def. 1, then no user u can conclude that any secret s is entailed by the KB she is interacting with—enhanced with the object-level background knowledge BK_u—for the following reasons: By point 1, KB and K have the same observable behavior, and K is a possible knowledge base for u since $K \in PKB_u$; therefore, as far as u knows, the knowledge base might be K. Moreover, by point 2, K and the object-level background knowledge BK_u do not suffice to entail the secret s.

In the rest of the paper we tacitly assume that no secret is violated a priori, that is, for all secrets $s \in S_u$ there exists $K \in PKB_u$ such that $s \notin Cn(K \cup BK_u)$.[4] *Moreover, in order to improve readability, we shall omit the user u from subscripts and argument lists* whenever u is irrelevant to the context.

The attacks discussed in Section 3 can be easily formalized in this setting; so, in general, the maximal secure views of SCM are not secure according to Def. 1.

Example 2. Example 1 can be formalized in our model as follows: The set of secrecies S is the set of all assertions $works\_for(e, d_n)$; $BK = \emptyset$ and PKB is the set of all the knowledge bases K that consist of assertions like (1) and (2), and such that for each axiom $Employee(e)$, K contains exactly one corresponding axiom $works\_for(e, d_i)$ and viceversa. The filtering function f maps each $K \in PKB$ on the maximal subset of K that entails none of S's members, that is, $f(K) = K \setminus S$ (by definition of PKB).

Note that f is injective over PKB, so condition 1 of Def. 1 is satisfied only if $K = KB$. So, if KB contains at least one secret, then the conditions of Def. 1 cannot be satisfied, that is, maximal secure SCM views are not secure in our model. Indeed, KB can be

[3] In practice, bk-models are finite, and filterings computable, but no such assumption will be technically needed until Sec. 7.

[4] Conversely, no filtering function can conceal a secret that is already known by the user.

reconstructed from the secure view by observing that $KB = f(KB) \cup \{works\_for(e, d_n) \mid Employee(e) \in f(KB) \wedge \forall i = 1, \ldots, n, works\_for(e, d_i) \notin f(KB)\}$. □

Similarly, the formalizations of the other attacks yield injective filtering functions (the details are left to the reader).

5 A meta-secure query answering mechanism

In this section we introduce a *secure filtering function*. It is formulated as an iterative process based on a *censor*, that is a boolean function that decides for each axiom whether it should be obfuscated to protect confidentiality. The iterative construction manipulates pairs $\langle X^+, X^- \rangle \in \wp(\mathcal{L}) \times \wp(\mathcal{L})$ that represent a meta constraint on possible knowledge bases: we say that a knowledge base K *satisfies* $\langle X^+, X^- \rangle$ iff K entails all the sentences in X^+ and none of those in X^- (formally, $Cn(K) \supseteq X^+$ and $Cn(K) \cap X^- = \emptyset$).

Let PAX (the set of *possible axioms*) be the set of axioms that may occur in the knowledge base according to the user's knowledge, i.e. $PAX = \bigcup_{K' \in PKB} K'$. Let $\nu = |PAX| + 1$ if PAX is finite and $\nu = \omega$ otherwise; let $\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots$ be any enumeration of PAX ($i < \nu$).[5] The secure view construction for a knowledge base K in a bk-model $\mathcal{M}$ consists of the following, inductively defined sequence of pairs $\langle K_i^+, K_i^- \rangle_{i \geq 0}$:

- $\langle K_0^+, K_0^- \rangle = \langle \emptyset, \emptyset \rangle$, and for all $1 \leq i < \nu$, $\langle K_{i+1}^+, K_{i+1}^- \rangle$ is defined as follows:
  - if $censor_{\mathcal{M}}(K_i^+, K_i^-, \alpha_{i+1}) = true$ then let $\langle K_{i+1}^+, K_{i+1}^- \rangle = \langle K_i^+, K_i^- \rangle$;
  - if $censor_{\mathcal{M}}(K_i^+, K_i^-, \alpha_{i+1}) = false$ and $K \models \alpha_{i+1}$ then $\langle K_{i+1}^+, K_{i+1}^- \rangle = \langle K_i^+ \cup \{\alpha_{i+1}\}, K_i^- \rangle$;
  - otherwise let $\langle K_{i+1}^+, K_{i+1}^- \rangle = \langle K_i^+, K_i^- \cup \{\alpha_{i+1}\} \rangle$.

Finally, let $K^+ = \bigcup_{i<\nu} K_i^+$, $K^- = \bigcup_{i<\nu} K_i^-$, and $f_{\mathcal{M}}(K, u) = K^+$.

Note that the inductive construction aims at finding maximal sets K^+ and K^- that (i) partly describe what does / does not follow from K (as K satisfies $\langle K^+, K^- \rangle$ by construction), and (ii) do not trigger the censor (the sentences α_{i+1} that trigger the censor are included neither in K^+ nor in K^-, cf. the induction step).

In order to define the censor we need an auxiliary definition that captures all the sentences that can be entailed with the background knowledge BK and the meta-knowledge PKB enriched by a given constraint $\langle X^+, X^- \rangle$ analogous to those adopted in the iterative construction: Let $Cn_{\mathcal{M}}(X^+, X^-)$ be the set of all axioms $\alpha \in \mathcal{L}$ such that

$$\text{for all } K' \in PKB \text{ such that } K' \text{ satisfies } \langle X^+, X^- \rangle, \alpha \in Cn(K' \cup BK) \,. \tag{3}$$

Now the censor is defined as follows: For all $X^+, X^- \subseteq \mathcal{L}$ and $\alpha \in \mathcal{L}$,

$$censor_{\mathcal{M}}(X^+, X^-, \alpha) = \begin{cases} true & \text{if there exists } s \in S \text{ s.t. either } s \in Cn_{\mathcal{M}}(X^+ \cup \{\alpha\}, X^-) \\ & \text{or } s \in Cn_{\mathcal{M}}(X^+, X^- \cup \{\alpha\}); \\ false & \text{otherwise.} \end{cases} \tag{4}$$

In other words, the censor checks whether telling either that α is derivable or that α is not derivable to a user aware that the knowledge base satisfies $\langle X^+, X^- \rangle$, restricts the

[5] We will show later how to restrict the construction to finite sequences, by approximating PAX.

set of possible knowledge bases enough to conclude that a secret s is entailed by the knowledge base and the background knowledge encoded by BK and PKB.

Note that the censor obfuscates α_{i+1} if *any* of its possible answers entail a secret, independently of the actual contents of K (the two possible answers "yes" and "no" correspond to conditions $s \in Cn_{\mathcal{M}}(X^+ \cup \{\alpha\}, X^-)$ and $s \in Cn_{\mathcal{M}}(X^+, X^- \cup \{\alpha\})$, respectively). In this way, roughly speaking, the knowledge bases that entail s are given the same observable behavior as those that don't. Under a suitable continuity assumption on $Cn_{\mathcal{M}}$, this enforces confidentiality:

Theorem 1. *If $Cn_{\mathcal{M}}(KB^+, KB^-) \subseteq \bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-)$, then $f_{\mathcal{M}}$ is secure w.r.t. $\mathcal{M}$.*

Proof. Let u and s be arbitrary members of U and S_u, respectively. We have to show that there exists a $K \in PKB_u$ satisfying the two conditions of Def. 1. Let $\langle KB_i^+, KB_i^- \rangle_{i<\nu}$ be the sequence underlying the construction of $f_{\mathcal{M}}(KB, u)$. By construction, for all $i < \nu$, $s \notin Cn_{\mathcal{M}}(KB_i^+, KB_i^-)$. Moreover, by the continuity hypothesis, $Cn_{\mathcal{M}}(KB^+, KB^-) \subseteq \bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-)$, where $KB^+ = \bigcup_{i<\nu} KB_i^+$ and $KB^- = \bigcup_{i<\nu} KB_i^-$. It follows that $s \notin Cn_{\mathcal{M}}(KB^+, KB^-)$. Then, by definition of $Cn_{\mathcal{M}}$, there exists $K \in PKB_u$ such that:

$$Cn(K) \supseteq KB^+ \quad (5) \qquad Cn(K) \cap KB^- = \emptyset \quad (6) \qquad s \notin Cn(K \cup BK_u)\,. \quad (7)$$

Since (7) is the second condition of Def. 1, we are only left to show the first one, that is, $f_{\mathcal{M}}(K, u) = f_{\mathcal{M}}(KB, u)$. It suffices to prove by induction on i that for all $i < \nu$,

$$\langle K_i^+, K_i^- \rangle = \langle KB_i^+, KB_i^- \rangle\,. \quad (8)$$

The base case is trivial. Induction step ($i > 0$): By induction hypothesis, (8) holds for $i-1$, therefore $censor_{\mathcal{M}}(K_{i-1}^+, K_{i-1}^-, \alpha_i) = censor_{\mathcal{M}}(KB_{i-1}^+, KB_{i-1}^-, \alpha_i)$. If the censors are true, then (8) follows directly from the induction hypothesis. If the censor is false, then α_i belongs to $KB^+ \cup KB^-$; note that K and KB agree on these formulae, by (5) and (6), so both knowledge bases insert α_i into the same element of the i-th pair and (8) holds. □

Examples of the behavior of $f_{\mathcal{M}}$ are deferred until Sec.7.

6 Approximating background knowledge

Of course, the actual confidentiality of a filtering $f(KB, u)$ depends on a careful definition of the user's background knowledge, that is, PKB_u and BK_u. If background knowledge is not exactly known, as it typically happens, then it can be safely approximated by *overestimating* it. More background knowledge means larger BK_u and smaller PKB_u, which leads to the following comparison relation $\leq_k$ over bk-models:

Definition 2. *Given two bk-models $\mathcal{M} = \langle KB, U, f, \langle S_u, PKB_u, BK_u \rangle_{u \in U} \rangle$ and $\mathcal{M}' = \langle KB', U', f', \langle S'_u, PKB'_u, BK'_u \rangle_{u \in U'} \rangle$, we write $\mathcal{M} \leq_k \mathcal{M}'$ iff*

1. *$KB = KB'$, $U = U'$, $f = f'$, and $S_u = S'_u$ (for all $u \in U$);*
2. *for all $u \in U$, $PKB_u \supseteq PKB'_u$ and $BK_u \subseteq BK'_u$.*

The next proposition proves that a bk-model $\mathcal{M}$ can be safely approximated by any $\mathcal{M}'$ such that $\mathcal{M} \leq_k \mathcal{M}'$:

Proposition 2. *If f is secure w.r.t. $\mathcal{M}'$ and $\mathcal{M} \leq_k \mathcal{M}'$, then f is secure w.r.t. $\mathcal{M}$.*

Consequently, a generic advice for estimating *BK* consists in including as many pieces of relevant knowledge as possible, for example:

(i) modelling as completely as possible the integrity constraints satisfied by the data, as well as role domain and range restrictions and disjointness constraints;

(ii) including in *BK* all the relevant public sources of formalized relevant knowledge (such as ontologies and triple stores).

While object-level background knowledge is dealt with in the literature, the general metaknowledge encoded by *PKB* is novel. Therefore, the next section is focussed on some concrete approximations of *PKB* and their properties.

7 Approximating and reasoning about possible knowledge bases

In this section, we investigate the real world situations where *the knowledge base KB is finite* and *so are all the components of bk-models* (U, S_u, BK_u, PKB_u); then we focus on PKB_u that contain only finite knowledge bases. Consequently, $f_{\mathcal{M}}$ will turn out to be decidable and we will study its complexity under different assumptions.

A language for defining *PKB* is a necessary prerequisite for the practical implementation of our framework and a detailed complexity analysis of $f_{\mathcal{M}}$. Here we express *PKB* as the set of all theories that are contained in a given set of *possible axioms PAX* [6] and satisfy a given, finite set *MR* of *metarules* like:

$$\alpha_1, \ldots, \alpha_n \Rightarrow \beta_1 \mid \ldots \mid \beta_m \quad (n \geq 0, m \geq 0)\,, \tag{9}$$

where all α_i and β_j are in $\mathcal{L}$ ($1 \leq i \leq n$, $1 \leq j \leq m$). Informally, (9) means that if *KB* entails $\alpha_1, \ldots, \alpha_n$ then *KB* entails also some of $\beta_1, \ldots, \beta_m$. Sets of similar metarules can be succintly specified using *metavariables*; they can be placed wherever individual constants may occur, that is, as arguments of assertions, and in nominals. A metarule with such variables abbreviates the set of its *ground instantiations*: Given a $K \subseteq \mathcal{L}$, let $ground_K(MR)$ be the ground (variable-free) instantiation of *MR* where metavariables are uniformly replaced by the individual constants occurring in K in all possible ways.

Example 3. Let $MR = \left\{ \exists R.\{X\} \Rightarrow A(X) \right\}$, where X is a metavariable, and let $K = \left\{ R(a,b) \right\}$. Then $ground_K(MR) = \left\{ (\exists R.\{a\} \Rightarrow A(a)),\ (\exists R.\{b\} \Rightarrow A(b)) \right\}$. □

If r denotes rule (9), then let $body(r) = \{\alpha_1, \ldots, \alpha_n\}$ and $head(r) = \{\beta_1, \ldots, \beta_m\}$. We say r is *Horn* if $|head(r)| \leq 1$. A set of axioms $K \subseteq \mathcal{L}$ *satisfies* a ground metarule r if either $body(r) \nsubseteq Cn(K)$ or $head(r) \cap Cn(K) \neq \emptyset$. In this case we write $K \models_m r$.

Example 4. Let A, B, C be concept names and R be a role name. The axiom set $K = \{A \sqsubseteq \exists R.B, A \sqsubseteq C\}$ satisfies $A \sqsubseteq \exists R \Rightarrow A \sqsubseteq B \mid A \sqsubseteq C$ but not $A \sqsubseteq \exists R \Rightarrow A \sqsubseteq B$. □

Moreover, if K satisfies all the metarules in $ground_K(MR)$ then we write $K \models_m MR$. Therefore the formal definition of *PKB* now becomes:

$$PKB = \{K \mid K \subseteq PAX \wedge K \models_m MR\}\,. \tag{10}$$

[6] Differently from Sec. 5, here *PKB* is defined in terms of *PAX*.

In accordance with Prop. 2, we approximate *PAX* in a conservative way. We will analyze two possible definitions:

1. $PAX_0 = KB$ (i.e., as a minimalistic choice we only assume that the axioms of *KB* are possible axioms; of course, by Prop. 2, this choice is safe also w.r.t. any larger *PAX* where *at least* the axioms of *KB* are regarded as possible axioms);
2. $PAX_1 = KB \cup \bigcup_{r \in ground_{KB}(MR)} head(r)$.

Remark 1. The latter definition is most natural if metarules are automatically extracted from *KB* with rule mining techniques, that typically construct rules using material from the given *KB* (then rule heads occur in *KB*).

Example 5. Consider again Example 1. The user's metaknowledge about *KB*'s completeness can be encoded with:

$$Employee(X) \Rightarrow works\_for(X, d_1) \mid \ldots \mid works\_for(X, d_n)\,, \tag{11}$$

where X is a metavariable. First let $PAX = PAX_1$. The secure view $f_{\mathcal{M}}(KB)$ depends on the enumeration order of *PAX*. If the role assertions $works\_for(e, d_i)$ precede the concept assertions $Employee(e)$, then, in a first stage, the sets KB^+_j are progressively filled with the role assertions with $d_i \neq d_n$ that belong to *KB*, while the sets KB^-_j accumulate all the role assertions that do not belong to *KB*. In a second stage, the sets KB^+_j are further extended with the concept assertions $Employee(e)$ such that e does not work for d_n. The role assertions $works\_for(e, d_n)$ of *KB* and the corresponding concept assertions $Employee(e)$ are neither in KB^+ nor in KB^-. Note that the final effect is equivalent to removing from *KB* all the axioms referring to the individuals that work for d_n. Analogously, in [7] the individuals belonging to a specified set are removed from all answers.

Next suppose that the role assertions $works\_for(e, d_i)$ follow the concept assertions $Employee(e)$, and that each $works\_for(e, d_i)$ follows all $works\_for(e, d_k)$ such that $k < i$. Now all the assertions $Employee(e)$ of *KB* enter KB^+, and all axioms $works\_for(e, d_i)$ with $i < n - 1$ enter either KB^+ or KB^-, depending on whether they are members of *KB* or not. Finally, the assertions $works\_for(e, d_i) \in Cn(KB)$ with $i \in \{n - 1, n\}$ are inserted neither in KB^+ nor in KB^-, because the corresponding instance of (11) with $X = e$ has the body in KB^+ and the first $n - 2$ alternatives in the head in KB^-, therefore a negative answer to $works\_for(e, d_{n-1})$ would entail the secret $works\_for(e, d_n)$ by (11). This triggers the censor for all assertions $works\_for(e, d_{n-1})$. Summarizing, with this enumeration ordering it is possible to return the complete list of employees; the members of d_n are protected by hiding also which employees belong to d_{n-1}.

Finally, let $PAX = PAX_0$. In this case, all possible knowledge bases are subsets of *KB*; the latter contains exactly one assertion $works\_for(e, d_{i(e)})$ for each employee e. Then, in order to satisfy (11), every $K \in PKB$ containing $Employee(e)$ must contain also $works\_for(e, d_{i(e)})$. It follows that $f_{\mathcal{M}}$ must remove all references to the individuals e that work for d_n, as it happens with the first enumeration of PAX_1. □

Definition 3. *A bk-model $\mathcal{M}$ is* canonical *if for all users $u \in U$, PAX_u is either PAX_0 or PAX_1 and PKB_u is defined by (10) for a given MR_u. Moreover, $\mathcal{M}$ is* in a description logic DL *if for all $u \in U$, all the axioms in KB, PKB_u, BK_u, and S_u belong to DL.*

The size of PAX_0 and PAX_1[7] is polynomial in the size of $KB \cup MR$, therefore PKB is finite and exponential in the size of $KB \cup MR$. Finiteness implies the continuity hypothesis on $Cn_{\mathcal{M}}$ of Theorem 1, and hence (using Theorem 1 and Prop. 2):

Theorem 2. *If $\mathcal{M}$ is canonical, then $f_{\mathcal{M}}$ is secure with respect to all $\mathcal{M}' \leq_k \mathcal{M}$.*

Proof. Since $\mathcal{M}$ is canonical, for all $u \in U$, PKB_u is finite and $\nu = |PAX_u| + 1 < \omega$. By construction, the sets KB_i^+ and KB_i^- in the sequence $\langle KB_i^+, KB_i^- \rangle_{i<\nu}$ grow monotonically with i, so $\bigcup_{i<\nu} KB_i^+ = KB_{\nu-1}^+$ and $\bigcup_{i<\nu} KB_i^- = KB_{\nu-1}^-$. Moreover, $Cn_{\mathcal{M}}$ is monotonic in both arguments, so $\bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-) = Cn_{\mathcal{M}}(KB_{\nu-1}^+, KB_{\nu-1}^-)$. It follows that

$$Cn_{\mathcal{M}}(\bigcup_{i<\nu} KB_i^+, \bigcup_{i<\nu} KB_i^-) = Cn_{\mathcal{M}}(KB_{\nu-1}^+, KB_{\nu-1}^-) = \bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-)\,,$$

that is, the continuity hypothesis of Theorem 1 is satisfied. Then, by Theorem 1, $f_{\mathcal{M}}$ is secure with respect to $\mathcal{M}$, and by Prop. 2, $f_{\mathcal{M}}$ is secure with respect to all $\mathcal{M}' \leq_k \mathcal{M}$. □

First we analyze the complexity of constructing the secure view $f_{\mathcal{M}}(KB)$ when the underlying description logic is tractable, like $\mathcal{EL}$ and DL-lite for example.

Lemma 1. *If the axioms occurring in MR and K are in a DL with tractable subsumption and instance checking, then checking $K \models_m MR$ is:*

1. *in P if either MR is ground or there exists a fixed bound on the number of distinct variables in MR;*
2. *coNP-complete otherwise.*

Proof. Point 1: $K \models_m MR$ can be checked as follows: For each $r \in ground_K(MR)$ and all axioms $\alpha \in body(r) \cup head(r)$ check whether $K \models_m r$ by verifying whether there exists either $\alpha \in body(r)$ such that $\alpha \notin Cn(K)$, or $\alpha \in head(r)$ such that $\alpha \in Cn(K)$. The cost of each test $K \models_m r$ is polynomial in the size of MR and K since membership in $Cn(K)$ is in P by hypothesis. The number of iterations is polynomial in the size of MR and K, too, because the hypothesis that the number of variables in r is bounded implies that $|ground_K(MR)|$ is polynomial in the size of MR and K.

Point 2: (Membership) The complementary test $K \not\models_m MR$ can be carried out in two steps: first guess an $r \in MR$ and a substitution σ that maps each metavariable in r on an individual constant occurring in K; second, check whether $K \models_m r\sigma$ does *not* hold. Checking $K \models_m r\sigma$ is in P (cf. point 1), so $K \not\models_m MR$ can be checked in nondetermistic polynomial time, and hence the original problem ($K \models_m MR$) is in coNP. Hardness follows by reducing to $K \not\models_m MR$ the clause subsumption problem: *Given two clauses (i.e. two sets of literals) G and H, is there a substitution σ such that $G\sigma \subseteq H$?* (if the answer is "yes" then *G subsumes H*). The problem is still NP-complete if all literals are positive, terms are function-free, and predicate arity is bounded by 2. Let $G = \{p_1, \ldots, p_n\}$ and H be two clauses satisfying these assumptions. Let $K = H$ (i.e. K is a set of assertions whose concept names and role names are the unary and binary predicates of H, respectively, and whose individual constants are the terms occurring in H). Let $MR = \{p_1, \ldots, p_n \Rightarrow\}$, where the terms occurring in G and not in H are regarded as metavariables. Now G subsumes H iff there exists a substitution σ such that $G\sigma \subseteq H$, iff there is an instance $r \in ground_K(MR)$ such that $K \not\models_m r$, iff $K \not\models_m MR$. □

[7] We assume here and in the following complexity results that axiom sets—and hence KBs—have a natural encoding as strings that determine their size.

With Lemma 1, one can prove the following two lemmas.

Lemma 2. *Let $\mathcal{M}$ range over canonical bk-models. If $\mathcal{M}$, s, X^+, and X^- are in a DL with tractable subsumption/instance checking, and the number of distinct variables in MR is bounded by a constant, then checking whether $s \in Cn_{\mathcal{M}}(X^+, X^-)$ is:*

1. *in P if MR is Horn and $PAX = PAX_1$;*
2. *coNP-complete if either MR is not Horn or $PAX = PAX_0$.*

Proof. Point 1: By standard logic programming techniques, a minimal $K \subseteq PAX$ satisfying MR and entailing X^+ can be obtained with the following PTIME construction:

$$K_0 = X^+, \quad K_{i+1} = K_i \cup \bigcup \{ head(r) \mid r \in ground_{K_i}(MR) \wedge body(r) \subseteq Cn(K_i) \}. \tag{12}$$

This sequence reaches its limit after at most $|PAX|$ iterations. Then $s \in Cn_{\mathcal{M}}(X^+, X^-)$ holds iff either $s \in K_{|PAX|}$ or $K_{|PAX|} \cap X^- \neq \emptyset$. Both tests are in P since $K_{|PAX|} \subseteq PAX$.

Point 2: Membership in coNP is straightforward ($s \notin Cn_{\mathcal{M}}(X^+, X^-)$ can be checked by guessing a $K \subseteq PAX$ that satisfies $\langle X^+, X^- \rangle$ and such that $s \notin Cn(K \cup BK)$). To prove hardness first assume that $PAX = PAX_0$. For each given 3-SAT instance, encode its n propositional variables and their negation with $2n$ concept names P_i and $\bar{P}_i$, respectively. Introduce a concept name C_k for each clause $c_k = l_{k,1} \vee l_{k,2} \vee l_{k,3}$. Let KB consist of all the inclusions $A \sqsubseteq P_i$ and $A \sqsubseteq \bar{P}_i$ ($1 \leq i \leq n$), plus all $L_{k,j} \sqsubseteq C_k$ s.t. $L_{k,j}$ is the encoding of l_{kj} ($j = 1, 2, 3$). Let $s = (A \sqsubseteq B)$, $BK = \emptyset$ and let MR consists of all the rules $(A \sqsubseteq P_i, A \sqsubseteq \bar{P}_i \Rightarrow)$, $(\Rightarrow L_{k,j} \sqsubseteq C_k)$, $(\Rightarrow A \sqsubseteq C_k)$. MR is Horn, and the given clause set is satisfiable iff there exists $K \subseteq KB = PAX_0$ such that $K \models_m MR$. For all such K, $s \notin Cn(K)$ because B does not occur in KB. Then the given clauses are satisfiable iff $s \notin Cn_{\mathcal{M}}(\emptyset, \emptyset)$. This proves that checking whether $s \in Cn_{\mathcal{M}}(X^+, X^-)$ is coNP-hard.

We are left to show a similar result under the assumption that MR is not Horn and $PAX = PAX_1$. Let KB, s, and BK be defined as before. Let MR be the set of all rules $(A \sqsubseteq P_i, A \sqsubseteq \bar{P}_i \Rightarrow)$, $(\Rightarrow A \sqsubseteq P_i \mid A \sqsubseteq \bar{P}_i)$, $(A \sqsubseteq \bar{L}_{k,1}, A \sqsubseteq \bar{L}_{k,2}, A \sqsubseteq \bar{L}_{k,3} \Rightarrow s)$, where each $\bar{L}_{kj}$ is the encoding of the complement of l_{kj}. Clearly, the given set of clauses is satisfied iff there exists $K \subseteq PAX_1$ such that $K \models_m MR$ and $s \notin Cn(K)$; this is equivalent to $s \notin Cn_{\mathcal{M}}(\emptyset, \emptyset)$. The theorem follows immediately. □

Lemma 3. *Let $\mathcal{M}$ be a canonical bk-model. If $\mathcal{M}$, s, X^+, and X^- are in a DL with tractable entailment problems, and there is no bound on the number of variables in the metarules of MR, then checking $s \in Cn_{\mathcal{M}}(X^+, X^-)$ is:*

1. *in P^{NP} if MR is Horn and $PAX = PAX_1$;*
2. *in Π_2^p if either MR is not Horn or $PAX = PAX_0$.*

Proof. To prove Point 1, we use the same algorithm used for Lemma 2.(1), based on the bottom-up construction defined by (12). However, due to the lack of bounds on metavariables, $ground_{K_i}(MR)$ can be exponentially large. Then the complexity of each iteration in (12) is determined with a different, nondeterministic algorithm: For each possible ground instance of a rule head (quadratically many due to arity bounds) use the NP oracle to guess an instance of the rule body and check (in polynomial time) whether it is entailed by K_i. The deterministic algorithm then runs in polynomial time using an NP oracle.

Point 2 can be proved with the naive nondeterministic algorithm that guesses a $K \subseteq PAX$ and checks whether (i) $K \in PKB$, (ii) $X^+ \subseteq Cn(K)$ and $X^- \cap Cn(K) = \emptyset$, and (iii) $s \in Cn(K \cup BK_u)$. Condition (i) can be verified by checking whether $K \models_m MR$; this test is NP-complete by Lemma 1.(2). Conditions (ii) and (iii) are in P by hypothesis. So the whole nondeterministic algorithm runs in polynomial time using an NP oracle. □

The value of $censor(X^+, X^-, \alpha)$ can be computed straightforwardly by iterating the tests $s \in Cn_{\mathcal{M}}(X^+ \cup \{\alpha\}, X^-)$ and $s \in Cn_{\mathcal{M}}(X^+, X^- \cup \{\alpha\})$ for all secrets $s \in S$. Since the set of secrets is part of the parameter $\mathcal{M}$ of the filtering function, the number of iterations is polynomial in the input and the complexity of the censor is dominated by the complexity of $Cn_{\mathcal{M}}()$. The latter is determined by Lemma 2 and Lemma 3, so we immediately get:

Corollary 1. *Let $\mathcal{M}$ be a canonical bk-model and suppose that $\mathcal{M}$, X^+, X^-, and α are in a DL with tractable entailment problems. If the number of distinct variables in MR is bounded by a constant, then computing $censor(X^+, X^-, \alpha)$ is:*

- *in P if MR is Horn and $PAX = PAX_1$;*
- *coNP-complete if either MR is not Horn or $PAX = PAX_0$.*

If there is no bound on the number of variables in the metarules of MR, then computing $censor(X^+, X^-, \alpha)$ is:

- *in P^{NP} if MR is Horn and $PAX = PAX_1$;*
- *in Π_2^p if either MR is not Horn or $PAX = PAX_0$.*

We are now ready to analyze the complexity of filtering functions:

Theorem 3. *If $\mathcal{M}$ is a canonical bk-model in a DL with tractable entailment problems, then computing $f_{\mathcal{M}}(KB)$ is:*

1. *in P if the number of distinct variables in the rules of MR is bounded, MR is Horn, and $PAX = PAX_1$;*
2. *P^{NP}-complete if the number of distinct variables in MR is bounded, and either MR is not Horn or $PAX = PAX_0$;*
3. *in P^{NP} if the variables in MR are unbounded, MR is Horn, and $PAX = PAX_1$;*
4. *in Δ_3^p if MR is not restricted and $PAX \in \{PAX_0, PAX_1\}$.*

Proof. Point 1 follows easily from Corollary 1: use the straightforward algorithm that iterates over all α_i in the enumeration of PAX, and for each of them computes the censor and checks whether $KB \models \alpha$ (if needed); since the number of iterations is polynomial in the input, the overall complexity is dominated by the complexity of evaluating the censor and KB entailments (both are tractable).

Point 2: Assume that $PAX = PAX_0$ and MR is Horn, the other case where $PAX = PAX_1$ and MR is not Horn can be proved with the techniques adopted in Lemma 2.(2). Membership in P^{NP} is straightforward, by the same argument applied in Point 1. Hardness is proved by a reduction of the maximum satisfying assignment problem which, given a set of clauses $C = \{c_1, \ldots, c_m\}$ in the variables $p_1, \ldots, p_n$, consists in finding the lexicographically maximum assignment $\mu^{msa} \in \{0,1\}^n$ that satisfies C, or 0 if C is unsatisfiable. We extend KB and MR defined at point 2 in Lemma 2 as follows: first, we add

to KB $A \sqsubseteq C$ and the set of inclusions $A \sqsubseteq P'_i$, with $1 \leq i \leq n$. Secondly, we replace the rules $\Rightarrow A \sqsubseteq C_k$, $1 \leq k \leq m$, with $A \sqsubseteq C \Rightarrow A \sqsubseteq C_k$ and add the rules $A \sqsubseteq P'_i \Rightarrow A \sqsubseteq P_i$, with $1 \leq i \leq n$. Finally, consider an ordering of PAX where $\alpha_1 = A \sqsubseteq C$ and, for each $1 \leq i \leq n$, $\alpha_{i+1} = A \sqsubseteq P'_{n+1-i}$ (the rest of the ordering is not relevant).

The inclusion $A \sqsubseteq C$ plays the role of a satisfiability checker for C, that is if C is not satisfiable, then for all $K \in KB$, $A \sqsubseteq C \notin Cn(K)$. Consequently, $A \sqsubseteq C \notin f_{\mathcal{M}}(KB)$. Assume now that C is satisfiable. First of all, since for all $0 \leq i \leq n$ $KB \models \alpha_i$, then $KB^- = \emptyset$ and $\alpha_i \in KB^+$ iff $censor_{\mathcal{M}}(KB_i^+, \emptyset, \alpha_i)$ is false. Now, note that the α_i are not forced to be entailed by any rule, therefore for each $K \in PKB$ also $K \setminus \{\alpha_i\} \in PKB$. Consequently, $censor_{\mathcal{M}}(KB_i^+, \emptyset, \alpha_i)$ is false iff there exists a $K \in PKB$ such that $Cn(K) \supseteq KB_i^+ \cup \{\alpha_i\}$. In particular, this ensures that $A \sqsubseteq C \in KB^+$. Now, since PKB satisfies the rules $A \sqsubseteq P'_i \Rightarrow A \sqsubseteq P_i$ and encodes with the inclusions $A \sqsubseteq P_i$ all possible assignments ν that satisfy C, this means that $censor_{\mathcal{M}}(KB_i^+, \emptyset, \alpha_i)$ is false (i.e. $\alpha_i \in KB^+$) iff there exists an assignment μ such that $\mu_i = 1$ and for all $1 \leq j < i$, $\mu_j = 1$ iff $A \sqsubseteq P'_j \in KB^+$. Finally, from the fact that most significant $A \sqsubseteq P'_i$ are processed first, we have that if C is not satisfiable then $A \sqsubseteq C \notin f_{\mathcal{M}}(KB)$, otherwise $A \sqsubseteq C \in f_{\mathcal{M}}(KB)$ and $A \sqsubseteq P'_i \in f_{\mathcal{M}}(KB)$ iff $\mu_i^{msa} = 1$.

Points 3, 4 are straightforward by the same argument for membership in Point 1. □

Theorem 4. *Computing $f_{\mathcal{M}}(KB)$ over canonical $\mathcal{M}$ in a DL with ExpTime entailment (e.g. $\mathcal{ALCQO}$, $\mathcal{ALCIO}$, $\mathcal{ALCQI}$, $\mathcal{SHOQ}$, $\mathcal{SHIO}$, $\mathcal{SHIQ}$), is still in ExpTime.*

Proof. Consider any test $s \in Cn_{\mathcal{M}}(X^+, X^-)$ in the construction of $f_{\mathcal{M}}(KB, u)$ (there are two such tests for each censor evaluation). Carrying out the test for given X^+, X^-, and $s \in S_u$ can be done by brute force, iterating over all the exponentially many $K' \subseteq PAX$ (which is either PAX_0 or PAX_1 whose size is polynomial in KB). For each such K', we have to verify whether it belongs to PKB, by checking whether $K' \models_m MR$; this can be done in ExpTime by iterating over all the (ground) instances $r \in ground_K(MR)$ and checking in polynomial time whether $K' \models_m r$. Then, for all $K' \in PKB$, three ExpTime problems must be solved ($X^+ \subseteq Cn(K')$, $X^- \cap Cn(K') = \emptyset$, and $s \notin Cn(K' \cup BK_u)$). If they all succeed, $s \notin Cn_{\mathcal{M}}(X^+, X^-)$; otherwise the algorithm continues with the next $K' \subseteq PAX$. So the total cost of each censor call is exponential in the size of KB, MR, and BK_u. In order to compute $f_{\mathcal{M}}(KB)$, this cost is iterated for all combinations of secrets and axioms in PAX; moreover, for each iteration where the censor is false, an additional ExpTime entailment problem is solved ($KB \models \alpha_{i+1}$). It follows that computing $f_{\mathcal{M}}(KB, u)$ is exponential in the size of KB, MR, BK_u, and S_u. □

Theorem 5. *Computing $f_{\mathcal{M}}(KB)$ over canonical $\mathcal{M}$ in $\mathcal{SROIQ}(\mathcal{D})$ is in $coNP^{N2ExpTime}$.*

Proof. (Hint) Use the same brute-force algorithm used in Theorem 4. □

8 Relationships with the SCM

Here we show that the meta-secure framework is a natural generalization of the SCM. The main result—roughly speaking—demonstrates that the SCM model can be essentially regarded as a special case of our framework where $PKB \supseteq \wp(KB)$ and $BK = \emptyset$. In this case $f_{\mathcal{M}}$ is secure even if $\mathcal{M}$ is not assumed to be canonical.

Theorem 6. *Let $\mathcal{M} = \langle KB, U, f_{\mathcal{M}}, \langle S_u, PKB_u, BK_u\rangle_{u\in U}\rangle$. If $PKB = \wp(KB)$, $BK = \emptyset$, and KB is finite, then*

1. $Cn_{\mathcal{M}}(KB^+, KB^-) = \bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-)$.
2. *For all enumerations of PAX, the corresponding $f_{\mathcal{M}}(KB, u)$ is logically equivalent to a maximal secure view KB_u of KB according to the SCM; conversely, for all maximal secure view KB_u of KB (according to the SCM) there exists an enumeration of PAX such that the resulting $f_{\mathcal{M}}(KB, u)$ is logically equivalent to KB_u.*
3. *$f_{\mathcal{M}}$ is secure w.r.t. $\mathcal{M}$ and w.r.t. any $\mathcal{M}' = \langle KB, U, f_{\mathcal{M}}, \langle S_u, PKB'_u, BK'_u\rangle_{u\in U}\rangle$ such that $PKB' \supseteq \wp(KB)$ and $BK' = \emptyset$.*

Proof. By the first hypothesis, $PAX = KB$. As a first consequence, for all $\alpha \in PAX$, $\alpha \in Cn(KB)$, and hence, by definition of the inductive sequence $\langle KB_i^+, KB_i^-\rangle_{i<\nu}$, we have that all for all $i < \nu$, $KB_i^- = \emptyset$. As a second consequence, for all $X^+ \subseteq KB$, we have $X^+ \in PKB = \wp(KB)$. Therefore X^+ is also the least $K \in PKB$ (up to logical equivalence) such that $Cn(K) \supseteq X^+$ and $Cn(K) \cap \emptyset = \emptyset$. This fact and the second hypothesis imply (by definition of $Cn_{\mathcal{M}}$) that

$$Cn_{\mathcal{M}}(X^+, \emptyset) = Cn(X^+)\,. \tag{13}$$

As a special case, we get $Cn_{\mathcal{M}}(KB_i^+, KB_i^-) = Cn(KB_i^+)$, for all $i < \nu$. Moreover, by compactness, $Cn(\bigcup_{i<\nu} KB_i^+) = \bigcup_{i<\nu} Cn(KB_i^+)$; then Point 1 follows by:

$$Cn_{\mathcal{M}}(KB^+, KB^-) = Cn(KB^+) = Cn(\bigcup_{i<\nu} KB_i^+) = \bigcup_{i<\nu} Cn(KB_i^+) = \bigcup_{i<\nu} Cn_{\mathcal{M}}(KB_i^+, KB_i^-).$$

Now let $\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots$ be any enumeration of PAX. By induction on i, it is easy to prove (using (13) and the definitions of K_{i+1}^+ and the censor) that for all $i < \nu$, $\alpha_i \notin K_i^+$ iff either $Cn(K_{i-1}^+ \cup \{\alpha_i\}) \cap S \neq \emptyset$ or $\alpha_i \in Cn(K_{i-1})$. It follows immediately that $\bigcup_{i<\nu} KB_i^+$ is logically equivalent to a maximal subset KB_u of KB that does not entail any secret. By definition, the same holds for $f_{\mathcal{M}}(KB, u) = \bigcup_{i<\nu} KB_i^+$.

Conversely, let KB_u be any maximal subset of KB that entails no secret, and let $n = |KB_u|$. Let $\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots$ be any enumeration of PAX such that $KB_u = \{\alpha_1, \ldots, \alpha_n\}$ (i.e. the sentences in KB_u precede those in $KB \setminus KB_u$). As in the above paragraph, it can be verified that for all $i < \nu$, $\alpha_i \notin K_{i+1}^+$ iff either $Cn(K_i^+ \cup \{\alpha_i\}) \cap S \neq \emptyset$ or $\alpha_i \in Cn(K_i)$. It follows that KB_n^+ is logically equivalent to KB_u, and for all $i > n$, $KB_i^+ = KB_n$. Consequently, $\bigcup_{i<\nu} KB_i^+$ is logically equivalent to KB_u, and so is $f_{\mathcal{M}}(KB, u) = \bigcup_{i<\nu} KB_i^+$. This completes the proof of Point 2.

Point 3: $f_{\mathcal{M}}$ is secure w.r.t. $\mathcal{M}$ by Theorem 1, whose hypothesis is satisfied by Point 1. It follows by Proposition 2, that $f_{\mathcal{M}}$ is also secure for all $\mathcal{M}'$ such that $\mathcal{M}' \leq_k \mathcal{M}$, which includes all $\mathcal{M}'$ that are identical to $\mathcal{M}$ with the exception of their possible knowledge bases PKB', and such that $PKB' \supseteq PKB = \wp(KB)$. □

Remark 2. Theorem 6 applies to every canonical $\mathcal{M}$ such that $MR = BK = \emptyset$, because $MR = \emptyset$ implies that $PAX_0 = PAX_1 = KB$ and hence $PKB = \wp(KB)$. This shows that the SCM can be regarded as a special case of our framework where the user has no background knowledge. Moreover, by this correpondence, one immediately obtains complexity bounds for the SCM from those for PAX_1 and Horn, bounded-variable MR.

9 Related work

Baader et al. [2], Eldora et al. [12], and Knechtel and Stuckenschmidt [14] attach security labels to axioms and users to determine which subset of the KB can be used by each subject. These works are instances of the SCM so they are potentially vulnerable to the attacks based on background knowledge; this holds in particular for [14] that pursues the construction of maximal secure views. Similar considerations hold for [16]. Moreover, in [2, 12] axiom labels are not derived from the set of secrets; knowledge engineers are responsible for checking ex post that no confidential knowledge is entailed; in case of leakage, the view can be modified with a revision tool based on pinpointing. Our mechanism produces automatically a secure view from the secrets, instead, and decides secondary protection, i.e. which additional axioms shall be hidden for security.

Chen and Stuckenschmidt [7] adopt an instance of the SACM based on removing some individuals entirely. In general, this may be secure against metaknowledge attacks (cf. Ex. 5). However, no methodology is provided for selecting the individuals to be removed given a target set of secrets.

In [3], KB is partioned into a visible part KB_v and a hidden part KB_h. Conceptually, this is analogous to axiom labelling, cf. the above approaches. Their confidentiality methodology seems to work only under the assumption that the signatures of KB_v and KB_h are disjoint, because in strong safety they do not consider the formulae that are implied by a combination of KB_v and KB_h. Surely the axioms of KB_h whose signature is included in the signature of KB_v cannot be protected, in general. A partition-based approach is taken in [10], too. It is not discussed how to select the hidden part KB_h given a set of target secrets (which includes the issue of deciding secondary protection).

Similarly, in [15] only ex-post confidentiality verification methods are provided. In their model the equivalent of PKB is the set of all knowledge bases that include a given set of publicly known axioms $\mathcal{S} \subseteq KB$; consequently, their verification method is vulnerable to the attacks to complete knowledge based on conditional metaknowledge (cf. Example 2 and Example 5) that cannot be encoded in their framework.

Cuenca Grau and Horrocks [9] investigate knowledge confidentiality from a probabilistic perspective: enlarging the public view should not change the probability distribution over the possible answers to a "sensitive query" Q that represents the set of secrets. In [9] users can query the knowledge base only through a pre-defined set of views (we place no such restriction, instead). A probability distribution P over the set of knowledge bases plays a role similar to metaknowledge. However, their confidentiality condition allows P to be replaced with a different P' after enlarging the public view, so at a closer look P does not really model the user's a priori knowledge about the knowledge base (that should remain constant), differently from our PKB.

Our method is inspired by the literature on *controlled (database) query evaluation* (CQE) based on lies and/or refusals ([4, 5, 6] etc). Technically we use *lies*, because rejected queries are not explicitly marked. However, our censor resembles the classical refusal censor, so the properties of $f_{\mathcal{M}}$ are not subsumed by any of the classical CQE methods. For example (unlike the CQE approaches that use lies), $f_{\mathcal{M}}(KB, u)$ encodes only correct knowledge, and it is secure even if users initially know a disjunction of secrets. Unlike the refusal method, $f_{\mathcal{M}}$ can handle *cover stories* because users are not

told which queries are obfuscated; as an additional advantage, our method needs not to adapt existing engines to handle nonstandard answers like refusals (*mum*).

10 Discussion and conclusions

We identified some novel vulnerabilities of those confidentiality preservation methods that do not take background knowledge into account. The new confidentiality model of Sec. 4 can detect these vulnerabilities, based on a generic formalization of object- and meta-level background knowledge. A general mechanism for constructing secure views (the filtering $f_{\mathcal{M}}$) is provably secure w.r.t. this model under a continuity assumption, and generalizes a few previous approaches (cf. Thm. 6 and Ex. 5). In order to compute secure views in practice we introduced a safe, generic method for approximating background knowledge, and a specific rule-based metalanguage. In this instantiation of the general framework $f_{\mathcal{M}}$ is always secure and its complexity can be analyzed.

If the underlying DL is tractable, then in the simplest case $f_{\mathcal{M}}$ can be computed in polynomial time. The number of variables in metarules and the adoption of a more secure approximation (PAX_0) may increase complexity up to $\mathrm{P}^{NP} = \Delta_2^p$ and perhaps Δ_3^p. The complexity of non-Horn metarules, however, can be avoided by replacing each non-Horn r with one of its Horn strengthenings: $body(r) \Rightarrow \alpha$ such that $\alpha \in head(r)$. This approximation is safe (because it restricts *PKB*), and opens the way to a systematic use of the low-complexity bk-models based on PAX_1 and Horn metarules.

For the many ExpTime-complete DL, secure view computation does not increase asymptotic complexity. So far, the best upper complexity bound for computing secure views in the description logic underlying OWL DL (i.e. $\mathcal{SROIQ}(\mathcal{D})$) is $\mathrm{coNP}^{N2ExpTime}$.

We plan to refine these complexity results and investigate different tradeoffs between information availability and computational complexity. Moreover, the idea of mining metarules from *KB* is particularly intriguing: it would be the first automated support to background knowledge approximation.

We are investigating implementations of the low-complexity frameworks (based on PAX_1 and Horn metarules) using the incremental engine versions available for Pellet and ELK to avoid repeated classifications in the iterative construction of $f_{\mathcal{M}}$. Metarule bodies can be evaluated with SPARQL. Answer set programming technologies (e.g. DLV-Hex [11]) provide interesting alternatives. Secure views are constructed off-line, so no overhead is placed on user queries, that can be answered with any standard engine. For these reasons, our approach is expected to be applicable in practice.

Bibliography

[1] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[2] F. Baader, M. Knechtel, and R. Peñaloza. A generic approach for large-scale ontological reasoning in the presence of access restrictions to the ontology's axioms. In *International Semantic Web Conference*, pages 49–64, 2009.

[3] J. Bao, G. Slutzki, and V. Honavar. Privacy-preserving reasoning on the semantic web. In *Web Intelligence*, pages 791–797. IEEE Computer Society, 2007.

[4] J. Biskup and P. A. Bonatti. Lying versus refusal for known potential secrets. *Data Knowl. Eng.*, 38(2):199–222, 2001.

[5] J. Biskup and P. A. Bonatti. Controlled query evaluation for enforcing confidentiality in complete information systems. *Int. J. Inf. Sec.*, 3(1):14–27, 2004.

[6] J. Biskup and P. A. Bonatti. Controlled query evaluation for known policies by combining lying and refusal. *Ann. Math. Artif. Intell.*, 40(1-2):37–62, 2004.

[7] W. Chen and H. Stuckenschmidt. A model-driven approach to enable access control for ontologies. In *Wirtschaftsinformatik*, volume 246 of *books@ocg.at*, pages 663–672. Österreichische Computer Gesellschaft, 2009.

[8] B. Cuenca Grau. Privacy in ontology-based information systems: A pending matter. *Semantic Web*, 1(1-2):137–141, 2010.

[9] B. Cuenca Grau and I. Horrocks. Privacy-preserving query answering in logic-based information systems. In Proc. of *ECAI'08*, pages 40–44. IOS Press, 2008.

[10] B. Cuenca Grau and B. Motik. Importing ontologies with hidden content. In *Description Logics*, volume 477 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

[11] T. Eiter, M. Fink, T. Krennwallner, and C. Redl. Conflict-driven ASP solving with external sources. *Theory and Practice of Logic Programming* 12(4-5):659-679, 2012.

[12] Eldora, M. Knechtel, and R. Peñaloza. Correcting access restrictions to a consequence more flexibly. In *Description Logics*, vol. 745 of *CEUR Workshop Proc*. CEUR-WS.org, 2011.

[13] P. Hitzler and T. Lukasiewicz, editors. *Web Reasoning and Rule Systems - 4th Int. Conference, RR 2010.*, volume 6333 of *Lecture Notes in Computer Science*. Springer, 2010.

[14] M. Knechtel and H. Stuckenschmidt. Query-based access control for ontologies. In Hitzler and Lukasiewicz [13], pages 73–87.

[15] P. Stouppa and T. Studer. Data privacy for knowledge bases. In *LFCS*, volume 5407 of *LNCS*, pages 409–421. Springer, 2009.

[16] J. Tao, G. Slutzki, and V. Honavar. Secrecy-preserving query answering for instance checking in $\mathcal{EL}$. In Hitzler and Lukasiewicz [13], pages 195–203.