TECHNISCHE UNIVERSITÄT DRESDEN

Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ Instance Stores (Extended Version)

Franz Baader, Francesco Kriegel, Adrian Nuradiansyah

LTCS-Report 19-01

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nöthnitzer Str. 46
Dresden

Contents

Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ Instance Stores (Extended Version)

Franz Baader, Francesco Kriegel, Adrian Nuradiansyah

Abstract

We make a first step towards adapting an existing approach for privacy-preserving publishing of linked data to Description Logic (DL) ontologies. We consider the case where both the knowledge about individuals and the privacy policies are expressed using concepts of the DL $\mathcal{EL}$, which corresponds to the setting where the ontology is an $\mathcal{EL}$ instance store. We introduce the notions of compliance of a concept with a policy and of safety of a concept for a policy, and show how optimal compliant (safe) generalizations of a given $\mathcal{EL}$ concept can be computed. In addition, we investigate the complexity of the optimality problem.

1 Introduction

When publishing information about individuals, one needs to ensure that certain privacy constraints are fulfilled. These constraints are encoded as *privacy policies*, and before publishing the information one needs to check whether the information is *compliant* with these policies [9]. We illustrate this setting using an example from [9]: when publishing information about hospitals, doctors, and patients, the policy may require that one should not be able to find out who are the cancer patients. In case the information to be published is not policy compliant, it first needs to be modified in a minimal way to make it compliant. However, compliance per se is not enough if a possible attacker can also obtain relevant information from other sources, which together with the published information might violate the privacy policy. *Safety* requires that the combination of the published information with any other compliant information is again compliant [9]. More information on privacy-preserving data publishing can be found in the survey [12].

In [9], this problem was investigated in a setting where the information to be published is given as a relational dataset with (labeled) null values, and the policy is given by a conjunctive query. In order to make a given dataset compliant or safe, one is basically allowed to replace constants (or null values) by new null

values. The paper investigates the complexity of deciding compliance (Is a given modification of a dataset policy compliant?), safety (Is a given modification of a dataset safe w.r.t. a policy?), and optimality (Is a given modification of a dataset safe w.r.t. a policy and does it change the dataset in a minimal way?). The obtained complexity results depend on whether combined or data complexity is considered, and whether closed- or open-world semantics are used. For combined complexity, they lie on the second and third level of the polynomial hierarchy. The paper does not consider the case where the information in the dataset is augmented by ontological knowledge. In [7], ontologies are used to formulate privacy policies, but the policies considered there are concerned with meta-information like location and duration of data storage, intended use of data, etc. In contrast, the policies considered in [9] and in the present paper specify what information needs to be hidden.

In the present paper, we make a first step towards handling ontologies in the context of privacy-preserving data publishing, but consider a quite restricted setting, where information about an individual is given by a concept of the inexpressive Description Logic (DL) $\mathcal{EL}$. Basically, this is the setting where the ontology consists of an ABox containing only concept assertions of the form $C(a)$ for possibly complex concepts C, but no role assertion. In [14], such an ABox was called an *instance store*. In addition, we assume that there is no TBox, i.e., all the information about the individual a is given by the concept C.[1] A policy is then given by an instance query, i.e., by an $\mathcal{EL}$ concept D. A concept C (giving information about some individual a) is *compliant* with this policy, if it is not subsumed by D, i.e., if $C(a)$ does not imply $D(a)$. In our example, the policy could be formalized as the $\mathcal{EL}$ concept

$$D = \mathit{Patient} \sqcap \exists \mathit{seen\_by}.(\mathit{Doctor} \sqcap \exists \mathit{works\_in}.\mathit{Oncology}),$$

which says that one should not be able to find out who are the patients that are seen by a doctor that works for the oncology department. The concept

$$C = \mathit{Patient} \sqcap \mathit{Male} \sqcap \exists \mathit{seen\_by}.(\mathit{Doctor} \sqcap \mathit{Female} \sqcap \exists \mathit{works\_in}.\mathit{Oncology})$$

is not compliant with the policy D since $C \sqsubseteq D$. The concept

$$C' = \mathit{Male} \sqcap \exists \mathit{seen\_by}.(\mathit{Doctor} \sqcap \mathit{Female} \sqcap \exists \mathit{works\_in}.\mathit{Oncology})$$

is a compliant generalization of C, i.e., $C \sqsubseteq C'$ and $C' \not\sqsubseteq D$. However, it is not safe since $C' \sqcap \mathit{Patient} \sqsubseteq D$, i.e., if the attacker already knows that a is a patient then together with $C'(a)$ the hidden information D is revealed. In contrast,

$$C'' = \mathit{Male} \sqcap \exists \mathit{seen\_by}.(\mathit{Doctor} \sqcap \mathit{Female} \sqcap \exists \mathit{works\_in}.\top),$$

[1] Since $\mathcal{EL}$ concepts are closed under conjunction, we can assume that the ABox contains only one assertion for each individual a.

is a safe generalization of C, though it is less obvious to see this. This concept is, however, not optimal since more information than necessary is removed. In fact, the concept

$$C''' = \mathit{Male} \sqcap \exists \mathit{seen\_by}.(\mathit{Doctor} \sqcap \mathit{Female} \sqcap \exists \mathit{works\_in}.\top) \sqcap \exists \mathit{seen\_by}.(\mathit{Female} \sqcap \exists \mathit{works\_in}.\mathit{Oncology})$$

is a safe generalization of C that is more specific than C'', i.e., $C \sqsubseteq C''' \sqsubset C''$.

We will show how to compute optimal compliant and optimal safe generalizations of $\mathcal{EL}$ concepts C with $\mathcal{EL}$ policies, but instead of only one policy concept we allow for a finite set of $\mathcal{EL}$ concepts as policy, where a concept C' is compliant with the policy $\{D_1, \ldots, D_p\}$ iff it is compliant with each element of this set, i.e., $C \not\sqsubseteq D_i$ holds for all $i = 1, \ldots, p$. In addition, following [9], we will also view optimality as a decision problem, and investigate its complexity. A short version of this paper, without the results of Section 5, was presented at DL 2018 [6].

2 Preliminaries

A wide range of DLs of different expressive power has been investigated in the literature [2]. Here, we only introduce the DL $\mathcal{EL}$, for which reasoning is tractable [4, 8, 1]. Let N_C and N_R be mutually disjoint sets of *concept* and *role names*, respectively. Then $\mathcal{EL}$ *concepts* over these names are constructed from concept names using the constructors top concept ($\top$), conjunction ($C \sqcap D$), and existential restriction ($\exists r.C$). The *size* of an $\mathcal{EL}$ concept C is the number of occurrences of $\top$ as well as concept and role names in C, its *role depth* is the maximal nesting of existential restrictions, and its *signature* $\mathit{sig}(C)$ is the set of all concept and role names occurring in C.

The semantics of $\mathcal{EL}$ is defined through *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set, called the *domain*, and $\cdot^{\mathcal{I}}$ is the *interpretation function*, which maps every $A \in N_C$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and every $r \in N_R$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. This function $\cdot^{\mathcal{I}}$ is extended to arbitrary $\mathcal{EL}$ concepts by setting $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$, $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$, and $(\exists r.C)^{\mathcal{I}} := \{\delta \in \Delta^{\mathcal{I}} \mid \exists \eta \in C^{\mathcal{I}}.(\delta, \eta) \in r^{\mathcal{I}}\}$.

The $\mathcal{EL}$ concept C is *subsumed by* the $\mathcal{EL}$ concept D (written $C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds for all interpretations $\mathcal{I}$. Strict subsumption (written $C \sqsubset D$) holds if $C \sqsubseteq D$ and $D \not\sqsubseteq C$, and we say that C is *equivalent* to D (written $C \equiv D$) if $C \sqsubseteq D$ and $D \sqsubseteq C$.

Subsumption between $\mathcal{EL}$ concepts can be decided in polynomial time. In [4], this was shown using a homomorphism characterization of subsumption, but it is also an easy consequence of the following result of Küsters. Given an $\mathcal{EL}$ concept C, we *reduce* it by exhaustively replacing subconcepts of the form $E \sqcap F$ with $E \sqsubseteq F$ by E (modulo associativity and commutativity of $\sqcap$). As shown in [16],

this can be done in polynomial time, and two concepts C, D are equivalent iff their reduced forms are equal up to associativity and commutativity of $\sqcap$.

We are now ready to define the important notions regarding privacy-preserving publishing of ontological information that will be investigated in this paper. As mentioned in the introduction, policies are finite sets of $\mathcal{EL}$ concepts. We assume in the following, that the concepts occurring in the policy are not equivalent to top since otherwise there would not be compliant concepts.

Definition 1. *A* policy *is a finite set $\mathcal{P} = \{D_1, \ldots, D_p\}$ of $\mathcal{EL}$ concepts such that $\top \not\equiv D_i$ for $i = 1, \ldots, p$. Given an $\mathcal{EL}$ concept C and a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$, the $\mathcal{EL}$ concept C' is*

- compliant *with $\mathcal{P}$ if $C' \not\sqsubseteq D_i$ holds for all $i = 1, \ldots, p$;*
- safe *for $\mathcal{P}$ if $C' \sqcap C''$ is compliant with $\mathcal{P}$ for all $\mathcal{EL}$ concepts C'' that are compliant with $\mathcal{P}$;*
- *a* $\mathcal{P}$-compliant generalization *of C if $C \sqsubseteq C'$ and C' is compliant with $\mathcal{P}$;*
- *an* optimal $\mathcal{P}$-compliant generalization *of C if it is a $\mathcal{P}$-compliant generalization of C and there is no $\mathcal{P}$-compliant generalization C'' of C such that $C'' \sqsubset C'$;*
- *a* $\mathcal{P}$-safe generalization *of C if $C \sqsubseteq C'$ and C' is safe for $\mathcal{P}$;*
- *an* optimal $\mathcal{P}$-safe generalization *of C if it is a $\mathcal{P}$-safe generalization of C and there is no $\mathcal{P}$-safe generalization C'' of C such that $C'' \sqsubset C'$.*

It is easy to see that safety implies compliance since the top concept is always compliant: if C' is safe for $\mathcal{P}$, then $\top \sqcap C' \equiv C'$ is compliant.

3 Computing optimal compliant generalizations

In this section, we characterize the concepts that are compliant with a given policy $\mathcal{P}$, and use this to develop an algorithm that computes all optimal $\mathcal{P}$-compliant generalizations of a given $\mathcal{EL}$ concept C.

But first, we recall the recursive characterization of subsumption in $\mathcal{EL}$ given in [5]. We call an $\mathcal{EL}$ concept an *atom* if it is a concept name or an existential restriction. Given an $\mathcal{EL}$ concept C, we denote the set of atoms occurring in its top-level conjunction with $\mathtt{con}(C)$. For example, if $C = A \sqcap \exists r.(B \sqcap \exists s.A)$, then $\mathtt{con}(C) = \{A, \exists r.(B \sqcap \exists s.A)\}$. Subsumption between atoms E, F can be characterized as follows: $E \sqsubseteq F$ iff

- $E = F \in N_C$ or

- there is $r \in N_R$ such that $E = \exists r.E', F = \exists r.F'$ and $E' \sqsubseteq F'$.

Definition 2. *Let S, T be sets of atoms. Then we say that S* covers *T if for every $F \in T$ there is $E \in S$ such that $E \sqsubseteq F$.*

With this notation, subsumption in $\mathcal{EL}$ can be characterized as follows.

Proposition 3. *Let C, D be $\mathcal{EL}$ concepts. Then $C \sqsubseteq D$ iff $\mathtt{con}(C)$ covers $\mathtt{con}(D)$.*

The following (polynomial-time decidable) characterization of compliance is an immediate consequence of this proposition.

Proposition 4. *The $\mathcal{EL}$ concept C' is compliant with the policy $\mathcal{P} = \{D_1, \ldots, D_p\}$ iff $\mathtt{con}(C')$ does not cover $\mathtt{con}(D_i)$ for any $i = 1, \ldots, p$, i.e., for every $i = 1, \ldots, p$, at least one of the following two properties holds:*

- *there is a concept name $A \in \mathtt{con}(D_i)$ such that $A \notin \mathtt{con}(C')$; or*
- *there is an existential restriction $\exists r.D \in \mathtt{con}(D_i)$ such that $C \not\sqsubseteq D$ for all existential restrictions of the form $\exists r.C \in \mathtt{con}(C')$.*

Now assume that we are given an $\mathcal{EL}$ concept C and a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$, and we want to construct a $\mathcal{P}$-compliant generalization C' of C. For C' to satisfy the condition of Proposition 4, there needs to exist for every $i = 1, \ldots, p$ an element of $\mathtt{con}(D_i)$ that is not covered by any element of $\mathtt{con}(C')$. In case $\mathtt{con}(C)$ contains elements covering such an atom, we need to remove or generalize them appropriately.

Definition 5. *We say that $H \subseteq \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$ is a* hitting set *of $\mathtt{con}(D_1), \ldots, \mathtt{con}(D_p)$ if $H \cap \mathtt{con}(D_i) \neq \emptyset$ for every $i = 1, \ldots, p$. This hitting set is* minimal *if there is no other hitting set strictly contained in it.*

Basically, the idea is now to choose a hitting set H of $\mathtt{con}(D_1), \ldots, \mathtt{con}(D_p)$ and use H to guide the construction of a compliant generalization of C. In order to make this generalization as specific as possible, we use minimal hitting sets. In case the policy contains concepts D_i with which C is already compliant (i.e., $C \not\sqsubseteq D_i$ holds), nothing needs to be done w.r.t. these concepts. This is why, in the following definition, $\mathtt{con}(D_i)$ does not take part in the construction of the hitting set if $C \not\sqsubseteq D_i$.

Definition 6. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. The set $SCG(C, \mathcal{P})$ of* specific compliant generalizations *of C w.r.t. $\mathcal{P}$ consists of the concepts that can be constructed from C as follows:*

- *If C is compliant with $\mathcal{P}$, then $SCG(C, \mathcal{P}) = \{C\}$.*

- *Otherwise, choose a minimal hitting set H of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$ where $i_1, \ldots, i_q$ are exactly the indices i for which $C \sqsubseteq D_i$. Note that $q \geq 1$ since we are in the case where C is not compliant with $\mathcal{P}$. In addition, according to our definition of a policy, none of the concepts D_i is equivalent to $\top$, and thus the sets $\mathtt{con}(D_{i_j})$ are non-empty. Consequently, at least one minimal hitting set exists. Each minimal hitting set H yields a concept in $SCG(C, \mathcal{P})$ by removing or modifying atoms in the top-level conjunction of C in the following way:*
  - *For every concept name $A \in \mathtt{con}(C)$, remove A from the top-level conjunction of C if $A \in H$;*
  - *For every existential restriction $\exists r_i.C_i \in \mathtt{con}(C)$, consider the set*

$$\mathcal{P}_i := \{G \mid \text{there is } \exists r_i.G \in H \text{ such that } C_i \sqsubseteq G\}.$$

    - *If $\mathcal{P}_i = \emptyset$, then leave $\exists r_i.C_i$ as it is.*
    - *If $\top \in \mathcal{P}_i$, then remove $\exists r_i.C_i$.*
    - *Otherwise, replace $\exists r_i.C_i$ with $\bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$.*

We can show that every element of $SCG(C, \mathcal{P})$ is a compliant generalization of C.

Proposition 7. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. If $C' \in SCG(C, \mathcal{P})$, then C' is a $\mathcal{P}$-compliant generalization of C.*

Proof. In case C is already compliant with $\mathcal{P}$, then $C = C'$ and we are done. Thus, assume that C is not compliant with $\mathcal{P}$. We show that C' is a compliant generalization of C by induction on the role depth of C.

First, we show that C' is a generalization of C, i.e., $C \sqsubseteq C'$. This is an easy consequence of the fact that, when constructing C' from C, atoms from the top-level conjunction of C are left unchanged, are removed, or are replaced by a conjunction of more general atoms. The only non-trivial case is where we replace an existential restriction $\exists r_i.C_i$ with the conjunction $\bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$. By induction, we know that $C_i \sqsubseteq F$ for all $F \in SCG(C_i, \mathcal{P}_i)$, and thus $\exists r_i.C_i \sqsubseteq \bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$.

Second, we show that C' is compliant with $\mathcal{P}$, i.e., $C' \not\sqsubseteq D_i$ holds for $i = 1, \ldots, p$. For the indices i with $C \not\sqsubseteq D_i$, we clearly also have $C' \not\sqsubseteq D_i$ since $C \sqsubseteq C'$. Now, consider one of the remaining indices $i_j \in \{i_1, \ldots, i_q\}$, where $i_1, \ldots, i_q$ are exactly the indices for which $C \sqsubseteq D_i$. The concept C' was constructed by taking some minimal hitting set H of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$. If the element in H hitting $\mathtt{con}(D_{i_j})$ is a concept name, then this concept name does not occur in $\mathtt{con}(C')$, and thus $C' \not\sqsubseteq D_{i_j}$. Thus, assume that it is an existential restriction $\exists r_i.G$. But then each existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ with $C_i \sqsubseteq G$ is

either removed or replaced by a conjunction of existential restrictions $\exists r_i.F$ such that (by induction) $F \not\sqsubseteq G$. In addition, other existential restrictions are either removed or generalized. This clearly implies $C' \not\sqsubseteq D_{i_j}$ since $\exists r_i.G$ in $\mathtt{con}(D_{i_j})$ is not covered by any element of $\mathtt{con}(C')$. □

However, $SCG(C, \mathcal{P})$ may also contain compliant generalizations of C that are not optimal, as illustrated by the following example.

Example 8. *Let* $C = \exists r.(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4)$ *and* $\mathcal{P} = \{D_1, D_2\}$*, where*

$$D_1 = \exists r.A_1 \sqcap \exists r.(A_2 \sqcap A_3) \quad \textit{and} \quad D_2 = \exists r.A_2 \sqcap \exists r.A_4.$$

We have $C \sqsubseteq D_1$ *and* $C \sqsubseteq D_2$*, and thus* C *is not compliant with* $\mathcal{P}$*. Consequently, the elements of* $SCG(C, \mathcal{P})$ *are obtained by considering the minimal hitting sets of* $\{\exists r.A_1, \exists r.(A_2 \sqcap A_3)\}$ *and* $\{\exists r.A_2, \exists r.A_4\}$*.*

If we take the minimal hitting set $H = \{\exists r.(A_2 \sqcap A_3), \exists r.A_2\}$ *and consider the only existential restriction in* $\mathtt{con}(C)$*, the corresponding set* $\mathcal{P}_i$ *consists of* $A_2 \sqcap A_3$ *and* A_2*. It is easy to see that* $SCG(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4, \mathcal{P}_i) = \{A_1 \sqcap A_3 \sqcap A_4\}$ *since the only minimal hitting set of* $\{A_1, A_2\}$ *and* $\{A_2\}$ *is* $\{A_2\}$*. Thus, we obtain* $C' := \exists r.(A_1 \sqcap A_3 \sqcap A_4)$ *as an element of* $SCG(C, \mathcal{P})$*.*

However, if we take the minimal hitting set $H' = \{\exists r.A_1, \exists r.A_2\}$ *instead, then the set* $\mathcal{P}'_i$ *corresponding to the only existential restriction in* $\mathtt{con}(C)$ *is* $\{A_1, A_2\}$*. Consequently, in this case* $SCG(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4, \mathcal{P}'_i) = \{A_3 \sqcap A_4\}$ *since the only minimal hitting set of* $\{A_1\}$ *and* $\{A_2\}$ *is* $\{A_1, A_2\}$*. This yields* $C'' := \exists r.(A_3 \sqcap A_4)$ *as another element of* $SCG(C, \mathcal{P})$*. Since* $C' \sqsubset C''$*, the element* C'' *cannot be optimal.*

The next lemma states that every compliant generalization of C subsumes some element of $SCG(C, \mathcal{P})$.

Lemma 9. *Let* C *be an* $\mathcal{EL}$*-concept and* $\mathcal{P} = \{D_1, \ldots, D_p\}$ *a policy. If* C'' *is a* $\mathcal{P}$*-compliant generalization of* C*, then there is* $C' \in SCG(C, \mathcal{P})$ *such that* $C' \sqsubseteq C''$*.*

Proof. If C is compliant with $\mathcal{P}$, then we have $C \in SCG(C, \mathcal{P})$ and $C \sqsubseteq C''$ since C'' is a generalization of C. Thus, assume that C is not compliant with $\mathcal{P}$, and let $i_1, \ldots, i_q$ be exactly the indices for which $C \sqsubseteq D_i$.

Now, let i_j be such an index. We have $C \sqsubseteq C'' \not\sqsubseteq D_{i_j}$ and $C \sqsubseteq D_{i_j}$. Since $C'' \not\sqsubseteq D_{i_j}$, there is an element $E_j \in \mathtt{con}(D_{i_j})$ that is not covered by any element of $\mathtt{con}(C'')$. Obviously, $H'' := \{E_1, \ldots, E_q\}$ is a hitting set of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$. Thus, there is a minimal hitting set H of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$ such that $H \subseteq H''$. Let C' be the element of $SCG(C, \mathcal{P})$ that was constructed using this hitting set H. We claim that $C' \sqsubseteq C''$. For this, it is sufficient to show that $\mathtt{con}(C')$ covers $\mathtt{con}(C'')$.

First, consider a concept name $A \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, we also have $A \in \mathtt{con}(C)$. If $A \notin H''$, then $A \notin H$, and thus A is not removed in the construction of C'. Consequently, $A \in \mathtt{con}(C')$ covers $A \in \mathtt{con}(C'')$. If $A \in H''$, then A is not covered by any element of $\mathtt{con}(C'')$ according to our definition of H'', which contradicts our assumption that $A \in \mathtt{con}(C'')$.

Second, consider an existential restriction $\exists r_i.E \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, there is an existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ such that $C_i \sqsubseteq E$. If this restriction is not removed or generalized when constructing C', then we are done since this restriction then belongs to $\mathtt{con}(C')$ and covers $\exists r_i.E$. Otherwise, $\mathcal{P}_i = \{G \mid \text{there is } \exists r_i.G \in H \text{ such that } C_i \sqsubseteq G\}$ is non-empty.

If $\top \in \mathcal{P}_i$, then $\exists r_i.\top \in H \subseteq H''$. However, then $\exists r_i.E \in \mathtt{con}(C'')$ covers an element of H'', which is a contradiction.

Consequently, $\top \notin \mathcal{P}_i$, and thus $\exists r_i.C_i$ is replaced with $\bigsqcap_{F \in SCG(C_i,\mathcal{P}_i)} \exists r_i.F$ when constructing C' from C. According to our definition of H'' and the fact that $H \subseteq H''$, none of the existential restrictions $\exists r_i.G$ considered in the definition of $\mathcal{P}_i$ is covered by $\exists r_i.E \in \mathtt{con}(C'')$. This implies that E is a $\mathcal{P}_i$-compliant generalization of C_i. By induction (on the role depth) we can thus assume that there is an $F \in SCG(C_i, \mathcal{P}_i)$ such that $F \sqsubseteq E$. This shows that $\exists r_i.E \in \mathtt{con}(C'')$ is covered by $\exists r_i.F \in \mathtt{con}(C')$. □

As an easy consequence of this lemma, we obtain that all optimal compliant generalizations of C must belong to $SCG(C, \mathcal{P})$.

Proposition 10. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. If C'' is an optimal $\mathcal{P}$-compliant generalization of C, then $C'' \in SCG(C, \mathcal{P})$ (up to equivalence of concepts).*

Proof. Let C'' be an optimal $\mathcal{P}$-compliant generalization of C. By Lemma 9, there is an element $C' \in SCG(C, \mathcal{P})$ such that $C' \sqsubseteq C''$. In addition, by Proposition 7, C' is a $\mathcal{P}$-compliant generalization of C. Thus, optimality of C'' implies $C'' \equiv C'$. □

We are now ready to formulate and prove the main result of this section.

Theorem 11. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. Then the set of all optimal $\mathcal{P}$-compliant generalizations of C can be computed in time exponential in the size of C and $D_1, \ldots, D_p$.*

Proof. It is sufficient to show that the set $SCG(C, \mathcal{P})$ can be computed in exponential time. In fact, given $SCG(C, \mathcal{P})$, we can compute the set of all optimal $\mathcal{P}$-compliant generalizations of C by removing elements that are not minimal w.r.t. subsumption, which requires at most exponentially many subsumption tests. Each subsumption test takes at most exponential time since subsumption

in $\mathcal{EL}$ is in P, and the elements of $SCG(C, \mathcal{P})$ have at most exponential size, as shown below.

We show by induction on the role depth that $SCG(C, \mathcal{P})$ consists of at most exponentially many elements of at most exponential size. The at most exponential cardinality of $SCG(C, \mathcal{P})$ is an immediate consequence of the fact that there are at most exponentially many hitting sets of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$, and each yields exactly one element of $SCG(C, \mathcal{P})$ (see Definition 6). Regarding the size of these elements, note that we may assume by induction that an existential restriction may be replaced by a conjunction of at most exponentially many existential restrictions, where each is of at most exponential size. The overall size of the concept description obtained this way is thus also of at most exponential size. Given this, it is easy to see that the computation of these elements also takes at most exponential time. □

The following example shows that the exponential upper bounds can indeed by reached.

Example 12. *Let* $C = P_1 \sqcap Q_1 \sqcap \ldots \sqcap P_n \sqcap Q_n$ *and* $\mathcal{P} = \{P_i \sqcap Q_i \mid 1 \leq i \leq n\}$. *Then* $SCG(C, \mathcal{P})$ *contains* 2^n *elements since the sets* $\{P_1, Q_1\}, \ldots, \{P_n, Q_n\}$ *obviously have exponentially many hitting sets. To be more precise,*

$$SCG(C, \mathcal{P}) = \{X_1 \sqcap \ldots \sqcap X_n \mid X_i \in \{P_i, Q_i\} \text{ for } i = 1, \ldots, n\}.$$

This example can easily be modified to enforce an element of exponential size. Consider $\widehat{C} = \exists r.C$ *and* $\widehat{\mathcal{P}} = \{\exists r.(P_i \sqcap Q_i) \mid 1 \leq i \leq n\}$. *Then* $SCG(\widehat{C}, \widehat{\mathcal{P}}) = \{\bigsqcap_{F \in SCG(C,\mathcal{P})} \exists r.F\}$. *We leave it to the reader to further modify the example in order to obtain exponentially many elements of exponential size.*

4 Computing optimal safe generalizations

Before we can characterize safety, we need to remove redundant elements from $\mathcal{P}$. We say that $D_i \in \mathcal{P}$ is *redundant* if there is a different element $D_j \in \mathcal{P}$ such that $D_i \sqsubseteq D_j$. The following lemma is easy to prove.

Lemma 13. *Let* $\mathcal{P}$ *be a policy and assume that* $D_i \in \mathcal{P}$ *is redundant. Then the following holds for all* $\mathcal{EL}$ *concepts* C, C'*:*

- C' *is compliant with* $\mathcal{P}$ *iff* C' *is compliant with* $\mathcal{P} \setminus \{D_i\}$*;*
- C *is safe for* $\mathcal{P}$ *iff* C *is safe for* $\mathcal{P} \setminus \{D_i\}$.

Proof. Obviously, compliance with $\mathcal{P}$ implies compliance with its subset $\mathcal{P} \setminus \{D_i\}$. Conversely, assume that C' is compliant with $\mathcal{P} \setminus \{D_i\}$. To show that C' is also

compliant with $\mathcal{P}$, we need to show that $C' \not\sqsubseteq D_i$. Thus, assume that $C' \sqsubseteq D_i$. Since D_i is redundant, there is $D_j \in \mathcal{P}$ different from D_i such that $D_i \sqsubseteq D_j$. But then $D_j \in \mathcal{P} \setminus \{D_i\}$, and thus $C' \sqsubseteq D_i \sqsubseteq D_j$ contradicts our assumption that C' is compliant with $\mathcal{P} \setminus \{D_i\}$.

The result for safety follows immediately from the result for compliance. □

This lemma shows that we can assume without loss of generality that our policies do not contain redundant concepts. However, elements of D_i of $\mathcal{P}$ may also contain redundant atoms. This can be avoided by reducing the policy concepts.

We call a policy *redundancy-free* if it does not contain redundant elements and every element is reduced.

The following proposition characterizes safety for redundancy-free policies.

Proposition 14. *Let $\mathcal{P} = \{D_1, \ldots, D_p\}$ be a redundancy-free policy. The $\mathcal{EL}$ concept C' is safe for $\mathcal{P}$ iff there is no pair of atoms (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and $E \sqsubseteq F$.*

Proof. First, assume that C' is not safe for $\mathcal{P}$, i.e., there is an $\mathcal{EL}$ concept C'' that is compliant with $\mathcal{P}$, but for which $C' \sqcap C''$ is not compliant with $\mathcal{P}$. The latter implies that there is $D_i \in \mathcal{P}$ such that $C' \sqcap C'' \sqsubseteq D_i$, which is equivalent to saying that $\mathtt{con}(C') \cup \mathtt{con}(C'')$ covers $\mathtt{con}(D_i)$. On the other hand, we know that $\mathtt{con}(C'')$ does not cover $\mathtt{con}(D_i)$ since C'' is compliant with $\mathcal{P}$. Thus, there is an element $F \in \mathtt{con}(D_i)$ that is covered by an element E of $\mathtt{con}(C')$. This yields (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and $E \sqsubseteq F$.

Conversely, assume that there is a pair of atoms (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_i)$, and $E \sqsubseteq F$. Let C'' be the concept obtained from D_i by removing F from the top-level conjunction of D_i. Then we clearly have $D_i \sqsubseteq C''$. In addition, since D_i is normalized, we also have $C'' \not\sqsubseteq D_i$. Consider $D_j \in \mathcal{P}$ different from D_i, and assume that $C'' \sqsubseteq D_j$. But then $D_i \sqsubseteq C'' \sqsubseteq D_j$ contradicts our assumption that $\mathcal{P}$ does not contain redundant elements. Thus, we have shown that C'' is compliant with $\mathcal{P}$. In addition, $\mathtt{con}(C') \cup \mathtt{con}(C'')$ covers $\mathtt{con}(D_i)$. In fact, the elements of $\mathtt{con}(D_i) \setminus \{F\}$ belong to $\mathtt{con}(C'')$, and thus cover themselves. In addition, F is covered by $E \in \mathtt{con}(C')$. Thus $C' \sqcap C'' \sqsubseteq D_i$, which shows that C' is not safe for $\mathcal{P}$. □

Clearly, the necessary and sufficient condition for safety stated in this proposition can be decided in polynomial time. If needed, the policy can first be made redundancy-free, which can also be done in polynomial time.

Corollary 15. *Safety of an $\mathcal{EL}$ concept for an $\mathcal{EL}$ policy is in P.*

We now consider the problem of computing optimal $\mathcal{P}$-safe generalizations of a given $\mathcal{EL}$ concept C. First note that, up to equivalence, there can be only one

optimal $\mathcal{P}$-safe generalization of C. This is an immediate consequence of the fact that the conjunction of safe concepts is again safe.

Lemma 16. *Let C'_1, C'_2 be two $\mathcal{EL}$ concepts that are $\mathcal{P}$-safe generalizations of C, where $\mathcal{P}$ is redundancy-free. Then $C'_1 \sqcap C'_2$ is also a $\mathcal{P}$-safe generalization of C.*

Proof. Clearly, $C \sqsubseteq C'_1$ and $C \sqsubseteq C'_2$ implies $C \sqsubseteq C'_1 \sqcap C'_2$. Regarding safety, assume that the condition of Proposition 14 is violated for $C'_1 \sqcap C'_2$, i.e., there is pair of atoms (E, F) such that $E \in \mathtt{con}(C'_1 \sqcap C'_2) = \mathtt{con}(C'_1) \cup \mathtt{con}(C'_2)$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_n)$, and $E \sqsubseteq F$. But then $E \in \mathtt{con}(C'_i)$ for $i = 1$ or $i = 2$, which implies that C'_1 or C'_2 does not satisfy the safety condition of Proposition 14. □

Thus there cannot be non-equivalent optimal $\mathcal{P}$-safe generalizations of a given $\mathcal{EL}$ concept C since their conjunction would then be more specific, contradicting their optimality. This property is independent of whether the policy is redundancy-free or not since turning a policy into one that is redundancy-free preserves the set of concepts that are compliant with (safe for) the policy.

Proposition 17. *If C'_1, C'_2 are optimal $\mathcal{P}$-safe generalizations of the $\mathcal{EL}$ concept C, then $C'_1 \equiv C'_2$.*

The following theorem shows how an optimal safe generalization of C can be constructed.

Theorem 18. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. We construct the concept C' from C by removing or modifying atoms in the top-level conjunction of C in the following way:*

- *For every concept name $A \in \mathtt{con}(C)$, remove A from the top-level conjunction of C if $A \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$;*
- *For every existential restriction $\exists r_i.C_i \in \mathtt{con}(C)$, consider the set of concepts*

  $$\mathcal{P}_i := \{G \mid \text{there is } \exists r_i.G \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ such that } C_i \sqsubseteq G\}.$$

  - *If $\mathcal{P}_i = \emptyset$, then leave $\exists r_i.C_i$ as it is.*
  - *If $\top \in \mathcal{P}_i$, then remove $\exists r_i.C_i$.*
  - *Otherwise, replace $\exists r_i.C_i$ with $\bigsqcap_{F \in OCG(C_i, \mathcal{P}_i)} \exists r_i.F$, where $OCG(C_i, \mathcal{P}_i)$ is the set of all optimal $\mathcal{P}_i$-compliant generalizations of C_i.*

Then C' is an optimal $\mathcal{P}$-safe generalization of C.

Proof. Obviously $C \sqsubseteq C'$ since, when constructing C' from C, atoms from the top-level conjunction of C are left unchanged, are removed, or are replaced by a conjunction of more general atoms.

To show that C' is safe for $\mathcal{P}$, we must show that the condition of Proposition 14 holds. Thus assume that it is violated, i.e., there is a pair of atoms (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and $E \sqsubseteq F$.

- First, we consider the case where $E = A$ is a concept name. Then $E \sqsubseteq F$ implies that $F = A$, and thus A is a concept name occurring in $\mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. However, all such concept names have been removed from the top-level conjunction of C when constructing C'. This contradicts our assumption that $E = A$ belongs to $\mathtt{con}(C')$.
- Second, assume that E is an existential restriction $E = \exists r_i.E'$. Then F is of the form $F = \exists r_i.G'$ and $E' \sqsubseteq G'$. In addition, there is an existential restriction $\exists r_i.C_i \in \mathtt{con}(C)$ from which $E = \exists r_i.E'$ was derived. By construction, $C_i \sqsubseteq E'$. In the construction of C', we consider the set $\mathcal{P}_i := \{G \mid$ there is $\exists r_i.G \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$ such that $C_i \sqsubseteq G\}$. Since $C_i \sqsubseteq E' \sqsubseteq G'$, this set is non-empty, and since $\exists r_i.E'$ is derived from $\exists r_i.C_i$, it does not contain $\top$. Consequently, we have $E' \in OCG(C_i, \mathcal{P}_i)$. However, $G' \in \mathcal{P}_i$ then implies that $E' \not\sqsubseteq G'$, which yields the desired contradiction.

It remains to show that C' is optimal. Thus assume that C'' is a $\mathcal{P}$-safe generalization of C. It is sufficient to show that $C' \sqsubseteq C''$, i.e., that $\mathtt{con}(C')$ covers $\mathtt{con}(C'')$.

- Assume that $A \in \mathtt{con}(C'')$ is a concept name. Then $C \sqsubseteq C''$ implies that $A \in \mathtt{con}(C)$. In addition, since C'' is safe for $\mathcal{P}$, Proposition 14 implies that $A \notin \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. Thus, A is not removed in the construction of C', which yields $A \in \mathtt{con}(C')$.
- Second, consider an existential restriction $\exists r_i.E \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, there is an existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ such that $C_i \sqsubseteq E$. If this restriction is not removed or generalized when constructing C', then we are done since this restriction then belongs to $\mathtt{con}(C')$ and covers $\exists r_i.E$. Otherwise, $\mathcal{P}_i = \{G \mid$ there is $\exists r_i.G \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$ such that $C_i \sqsubseteq G\}$ is non-empty.

  If $\top \in \mathcal{P}_i$, then $\exists r_i.\top \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. However, then $\exists r_i.E \in \mathtt{con}(C'')$ covers an element of $\mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, which is a contradiction to our assumption that C'' is safe for $\mathcal{P}$.

  Consequently, $\top \notin \mathcal{P}_i$, and thus $\exists r_i.C_i$ is replaced with $\bigsqcap_{F \in OCG(C_i, \mathcal{P}_i)} \exists r_i.F$ when constructing C' from C. Since C'' is safe for $\mathcal{P}$, none of the existential restrictions $\exists r_i.G$ considered in the definition of $\mathcal{P}_i$ is covered by

$\exists r_i.E \in \mathtt{con}(C'')$. This implies that E is a $\mathcal{P}_i$-compliant generalization of C_i. Consequently, there is an $F \in OCG(C_i, \mathcal{P}_i)$ such that $F \sqsubseteq E$. This shows that $\exists r_i.E \in \mathtt{con}(C'')$ is covered by $\exists r_i.F \in \mathtt{con}(C')$.

□

Since, by Theorem 11, $OCG(C_i, \mathcal{P}_i)$ can be computed in exponential time, the construction described in Theorem 18 can also be performed in exponential time.

Corollary 19. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. Then an optimal $\mathcal{P}$-safe generalization of C can be computed in exponential time.*

Example 12 can easily be modified to provide an example that shows that this exponential bound can actually not be improved since there are cases where the safe generalization is of exponential size.

5 The complexity of deciding optimality

In this section, we consider *optimality as a decision problem*, i.e., given $\mathcal{EL}$ concepts C, C' such that $C \sqsubseteq C'$ and a policy $\mathcal{P}$, decide whether C' is an optimal $\mathcal{P}$-compliant ($\mathcal{P}$-safe) generalization of C.

Theorem 11 and Corollary 19 show that the optimality problem is *in ExpTime* both for compliance and for safety. In fact, according to Theorem 11, given C and $\mathcal{P}$, we can compute the set of all optimal $\mathcal{P}$-compliant generalizations of C (up to equivalence) in exponential time. Consequently, this set contains at most exponentially many elements and each element has at most exponential size. This implies that we can test, in exponential time, whether a give concept C' is equivalent to one of the elements of this set. If this is the case, then C' is an optimal $\mathcal{P}$-compliant generalization of C, and otherwise not. The case of safety can be treated similarly, using Corollary 19 instead of Theorem 11.

In the following, we show that this complexity upper bound can be improved to coNP. Actually, we will prove this upper bound not just for compliance and safety, but for a whole class of properties.

Definition 20. *Let F be a function that assigns a set of $\mathcal{EL}$ concepts to every input consisting of an $\mathcal{EL}$ concept C and a policy $\mathcal{P}$. We say that the function F defines a* polynomial, upward-closed property *if the following holds for every input $C, \mathcal{P}$:*

- *for every $\mathcal{EL}$ concept C', we can decide $C' \in F(C, \mathcal{P})$ in time polynomial in $C, C', \mathcal{P}$ (polynomiality);*

- *if* $C' \in F(C, \mathcal{P})$ *and* $C' \sqsubseteq C''$*, then* $C'' \in F(C, \mathcal{P})$ *(upward-closedness).*

We say that C' *is an* optimal F-generalization of C w.r.t. $\mathcal{P}$ *if* $C \sqsubseteq C'$*,* $C' \in F(C, \mathcal{P})$*, and there is no* $C \sqsubseteq C'' \sqsubset C'$ *such that* $C'' \in F(C, \mathcal{P})$*.*

It is easy to see that compliance and safety are polynomial, upward-closed properties. In fact, upward-closedness is an obvious consequence of the definition of compliance (safety). For compliance, polynomiality follows from the fact that subsumption in $\mathcal{EL}$ can be decided in polynomial time. For safety, it is stated in Corollary 15. In addition, the notion of optimality introduced in the above definition coincides with the notion of optimality introduced in Definition 1 for compliance and safety.

We will show that, for polynomial, upward-closed properties, the optimality problem is in coNP, i.e., there is an *NP-algorithm* that, on input $C \sqsubseteq C'$ and $\mathcal{P}$, succeeds iff C' is *not* an optimal F-generalization of C w.r.t. $\mathcal{P}$. Basically, this algorithm proceeds as follows. It guesses a lower neighbor C'' of C' subsuming C, i.e., a concept C'' such that (i) $C \sqsubseteq C'' \sqsubset C'$ and (ii) there is no concept C''' with $C'' \sqsubset C''' \sqsubset C'$. If $C'' \in F(C, \mathcal{P})$, then the algorithm succeeds, and otherwise it fails.

To make this algorithm more concrete, we need to investigate the strict subsumption relation $\sqsubset$ on $\mathcal{EL}$ concepts in more detail. Following [3], we define the *one-step relation* $\sqsubset_1$ induced by $\sqsubset$ as

$$\sqsubset_1 := \{(C'', C') \in \sqsubset \mid \text{there is no } C''' \text{ such that } C'' \sqsubset C''' \sqsubset C'\}.$$

If $C'' \sqsubset_1 C'$ then we call C' an *upper neighbor* of C'' and C'' a *lower neighbor* of C'. In [3] it was shown that the relation $\sqsubset$ on $\mathcal{EL}$ concepts is *one-step generated*, i.e., the transitive closure of $\sqsubset_1$ is again $\sqsubset$. In the context of the optimality problem for polynomial, upward-closed properties, this implies the following: whenever there is a counterexample to the optimality of C' (i.e., a concept C'' such that $C \sqsubseteq C'' \sqsubset C'$ and $C'' \in F(C, \mathcal{P})$), then there is a lower neighbor of C' that provides such a counterexample. To see this, just note that $C'' \sqsubset C'$ implies that C' can be reached by a $\sqsubset_1$-chain from C''. The last element in this chain before C' is a lower neighbor of C', and it belongs to $F(C, \mathcal{P})$ since F is upward-closed.

Another interesting result in [3] is the following characterization of upper neighbors: for a given *reduced* $\mathcal{EL}$ concept C, the set of upper neighbors of C consists (up to equivalence) of the concepts D obtained from C as follows:

- Remove a concept name A from the top-level conjunction of C.
- Remove an existential restriction $\exists r.E$ from the top-level conjunction of C, and replace it by the conjunction of all existential restrictions $\exists r.F$ where F ranges over all upper neighbors of E.

Note that a special case of the second item is the removal of an existential restriction of the form $\exists r.\top$ since $\top$ does not have any upper neighbors. As shown in [15], this characterization implies that a given concept has only polynomially many upper neighbors, each of which is of polynomial size. As an easy consequence, we obtain the following lemma:

Lemma 21. *The one-step relation $\sqsubset_1$ induced by $\sqsubset$ on $\mathcal{EL}$ concepts is decidable in polynomial time.*

Regarding lower neighbors, it is sufficient for our purposes to show that they can be guessed in non-deterministic polynomial time. Thus, we are looking for an NP-algorithm that, given input concepts $C \sqsubseteq C'$, generates exactly the lower neighbors of C' that subsume C. Below, we sketch how an appropriate NP-algorithm can be obtained. A more detailed description as well as *proofs can be found in [15].* First, note that the lower neighbors C'' of C' can be obtained by conjoining an atom not implied by C' to C'. In addition, $C \sqsubseteq C''$ implies that $sig(C'') \subseteq sig(C)$. Given an $\mathcal{EL}$ concept C' and a finite set Σ of concept and role names, the set of *lowering atoms* for C' w.r.t. Σ is defined as

$$\begin{aligned} LA_\Sigma(C') \;:=\; & \{A \in \Sigma \cap N_C \mid A \notin \mathtt{con}(C')\} \cup \{\exists r.D \mid r \in \Sigma \cap N_R, sig(D) \subseteq \Sigma, \\ & C' \not\sqsubseteq \exists r.D, \text{ and } C' \sqsubseteq \exists r.E \text{ for all } E \text{ with } D \sqsubset_1 E\}. \end{aligned}$$

Lemma 22. *Let C' be an $\mathcal{EL}$ concept and Σ a finite set of concept and role names with $sig(C') \subseteq \Sigma$. Then C'' is a lower neighbor of C' with $sig(C'') \subseteq \Sigma$ iff there is an atom $At \in LA_\Sigma(C')$ such that $C'' \equiv C' \sqcap At$.*

Intuitively, adding a single atom to the top-level conjunction of C' is sufficient to obtain a lower neighbor since adding two (non-redundant) atoms would step too far down in the subsumption hierarchy. The same is true for adding an existential restriction $\exists r.D$ for which $\exists r.E$ with $D \sqsubset_1 E$ does not subsume C' since then $C' \sqcap \exists r.D \sqsubset C' \sqcap \exists r.E \sqsubset C'$ would hold.

Example 23. *Let $\Sigma := \{r, A_1, A_2, B_1, B_2, C_1, C_2\}$ and*

$$C' := \exists r.(A_1 \sqcap A_2 \sqcap B_1 \sqcap B_2) \sqcap \exists r.(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2) \sqcap \exists r.(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2).$$

Then, for all $i, j, k \in \{1, 2\}$, the existential restriction $\exists r.D$ with $D := A_i \sqcap B_j \sqcap C_k$ belongs to $LA_\Sigma(C')$. In fact, $C' \not\sqsubseteq \exists r.D$ is obviously true, and since the upper neighbors of D are $A_i \sqcap B_j$, $B_j \sqcap C_k$, and $A_i \sqcap C_k$, we also have $C' \sqsubseteq \exists r.E$ for all E with $D \sqsubset_1 E$. Obviously, by using n instead of three pairs of concept names, we can produce a generalized version of this example that shows that the cardinality of $LA_\Sigma(C')$ can be exponential in the size of C' and Σ.

In order to obtain an NP-algorithm that generates exactly the lower neighbors of C' that subsume C, it is sufficient to generate all lowering atoms for C' w.r.t.

$\Sigma := sig(C)$, and then remove the ones that do not subsume C. Unfortunately, the definition of lowering atoms given above Lemma 22 does not tell us directly how appropriate existential restrictions $\exists r.D$ can be found. The following necessary conditions follows from the characterization of lower neighbors given in [15].

Lemma 24. *Let C' be reduced. If $\exists r.D \in LA_\Sigma(C')$, then there is a set of existential restrictions $\{\exists r.F'_1, \ldots, \exists r.F'_k\} \subseteq \mathtt{con}(C')$ and $F_1 \in LA_\Sigma(F'_1), \ldots, F_k \in LA_\Sigma(F'_k)$ such that $D \equiv F_1 \sqcap \ldots \sqcap F_k$.*

We illustrate this lemma using the lowering atom $D = A_i \sqcap B_j \sqcap C_k$ in Example 23. Here we take the set of all existential restrictions in $\mathtt{con}(C')$ and choose $C_k \in LA_\Sigma(A_1 \sqcap A_2 \sqcap B_1 \sqcap B_2)$, $B_j \in LA_\Sigma(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2)$, and $A_i \in LA_\Sigma(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2)$. Obviously, D is indeed equivalent to the conjunction of these three atoms.

In general, not all choices of subsets and lower neighbors yields an appropriate existential restriction. For instance, if we take a smaller set of existential restrictions in our example (e.g., $\{\exists r.(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2), \exists r.(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2)\}$), then the obtained conjunction of lowering atoms (e.g., $B_1 \sqcap A_2$) is not appropriate since the corresponding existential restriction (e.g., $\exists r.(B_1 \sqcap A_2)$) is subsumed by C'.

The *NP-algorithm* generating exactly the elements of $LA_\Sigma(C')$ works as follows: given a reduced concept C' and a finite set Σ of concept and role names such that $sig(C') \subseteq \Sigma$, it non-deterministically chooses one of the following two alternatives:

1. Choose a concept name $A \in \Sigma \setminus \mathtt{con}(C')$, and output A. If there is no such concept name, fail.
2. Choose $r \in \Sigma \cap N_R$, a set of existential restrictions $\{\exists r.F'_1, \ldots, \exists r.F'_k\} \subseteq \mathtt{con}(C')$, and recursively guess elements $F_1 \in LA_\Sigma(F'_1), \ldots, F_k \in LA_\Sigma(F'_k)$. If for some $i, 1 \leq i \leq k$, the attempt to produce the atom $F_i \in LA_\Sigma(F'_i)$ fails, or if $C' \sqsubseteq \exists r.(F_1 \sqcap \ldots \sqcap F_k)$, or if $F_1 \sqcap \ldots \sqcap F_k$ has an upper neighbor E such that $C' \not\sqsubseteq \exists r.E$, then fail. Otherwise, output $\exists r.(F_1 \sqcap \ldots \sqcap F_k)$.

Lemma 25. *The algorithm described above runs in non-deterministic polynomial time, and its non-failing runs produce exactly the elements of $LA_\Sigma(C')$.*

Proof. Soundness of the algorithm is an immediate consequence of the fact that, in the second case, we explicitly test whether the conditions in the definition of lowering atoms are satisfied. Completeness is an easy consequence of Lemma 24. Finally, the choice of a concept name, a role name, and a subset of the existential restrictions in $\mathtt{con}(C')$, can clearly be achieved by making polynomially many binary choices. By induction on the role depth, we can assume that the algorithm can produce the elements $F_i \in LA_\Sigma(F'_i)$ in non-deterministic polynomial time, which shows that the overall algorithm runs in non-deterministic polynomial time. □

With this lemma in place, we can now show that the optimality problem for polynomial, upward-closed properties is in coNP.

Theorem 26. *Let F be a polynomial, upward-closed property. The problem of deciding, for a given input $C, C', \mathcal{P}$, whether C' is an optimal F-generalization of C w.r.t. $\mathcal{P}$ is in coNP.*

Proof. We show that non-optimality can be decided by an NP-algorithm, i.e., we describe an NP-algorithm that, given $C, C', \mathcal{P}$, succeeds iff C' is *not* an optimal F-generalization of C w.r.t. $\mathcal{P}$.

1. Check whether $C \sqsubseteq C'$ and $C' \in F(C, \mathcal{P})$. If this is not the case, then succeed. Otherwise, continue with the next step. Polynomiality of F and of subsumption in $\mathcal{EL}$ implies that this test can be done in polynomial time.
2. Set $\Sigma := sig(C)$ and guess a lowering atom $At \in LA_\Sigma(C')$. If $C \not\sqsubseteq At$, then fail. Otherwise, we know that $C'' := C' \sqcap At$ is a lower neighbor of C' that subsumes C, and we continue with the next step. As shown above, the elements of $LA_\Sigma(C')$ can be generated by an NP-algorithm.
3. Check whether $C'' \in F(C, \mathcal{P})$. If this is the case, then succeed, and otherwise fail.

It is easy to see that this algorithm is correct and runs in non-deterministic polynomial time. □

Since compliance and safety are polynomial, upward-closed properties, the following corollary is an immediate consequence of this theorem.

Corollary 27. *The optimality problem is in coNP for compliance and for safety.*

At the moment, we do not know whether these problems are also coNP-hard. We can show, however, that the Hypergraph Duality Problem [10] can be reduced to them. Note that this problem is in coNP, but conjectured to be neither in P nor coNP-hard [11, 13]. Given two finite families of inclusion-incomparable sets $\mathcal{G}$ and $\mathcal{H}$, the *Hypergraph Duality Problem* (DUAL) asks whether $\mathcal{H}$ consists exactly of the minimal hitting sets of $\mathcal{G}$.

Proposition 28. *There is a polynomial reduction of* DUAL *to the optimality problem that works both for compliance and for safety.*

Proof. Let $\mathcal{G} = \{G_1, \ldots, G_g\}, \mathcal{H} = \{H_1, \ldots, H_h\}$ be finite families of inclusion-incomparable sets and $G := G_1 \cup \ldots \cup G_g$. Since it can be checked in polynomial time whether a given set H is a minimal hitting set of $\mathcal{G}$, we can assume without loss of generality that all sets H_i are indeed minimal hitting sets of $\mathcal{G}$. The

problem to be decided by our reduction is thus whether $\mathcal{H}$ really contains *all* minimal hitting sets of $\mathcal{G}$. We view the elements of G as concept names, for $S \subseteq G$ write $\bigsqcap S$ for the conjunction of the concept names in S, and define

- $C := \exists r_1.\bigsqcap G$ and $\mathcal{P} := \{D_1 := \exists r_1.\bigsqcap G_1, \ldots, D_g := \exists r_1.\bigsqcap G_g\}$;
- $C' := \exists r_1.\bigsqcap(G \setminus H_1) \sqcap \ldots \sqcap \exists r_1.\bigsqcap(G \setminus H_h)$.

It is easy to see that C' is a $\mathcal{P}$-compliant and $\mathcal{P}$-safe generalization of C.

According to Definition 6 and the proof of Theorem 11, C has exactly one optimal $\mathcal{P}$-compliant generalization, which is obtained as follows. First, note that the top-level conjunctions of C and $D_1, \ldots, D_g$ respectively consist of a single existential restriction for the same role r_1, and that the concepts D_i are pairwise incomparable. This implies that on this level only one hitting set is considered, which is $\mathcal{P}$. On the next role level, we have $\mathcal{P}_1 = \{\bigsqcap G_1, \ldots, \bigsqcap G_g\}$. The optimal $\mathcal{P}_1$-compliant generalizations of $C_1 := \bigsqcap G$ are obtained by considering all minimal hitting sets of $G_1, \ldots, G_g$, and removing their elements from the top-level conjunction of C_1. Consequently, the optimal $\mathcal{P}$-compliant generalization of C is given as

$$C'' := \bigsqcap_{H \text{ minimal hitting set of } \mathcal{G}} \exists r_1.\bigsqcap(G \setminus H).$$

A close look at Theorem 18 reveals that C'' is also the optimal $\mathcal{P}$-safe generalization of C. This shows that C' is optimal for compliance (safety) iff $\mathcal{H}$ contains all minimal hitting sets of $\mathcal{G}$. □

6 Conclusion

We have introduced the notions of compliance with and safety for a policy in the simple setting where both the knowledge about individuals and the policy are given by $\mathcal{EL}$ concepts. In this setting, we were able to characterize compliant (safe) generalization of a given concept w.r.t. a policy, and have used these characterizations to obtain algorithms for computing optimal generalizations. These algorithms need exponential time, which is optimal since the generalizations may be of exponential size. For the optimality problems, we have provided a coNP upper bound, and have shown by a reduction from DUAL that they are unlikely to be in P since this would show DUAL $\in$ P, a problem that has been open for a long time.

In the future, we intend to extend this work in two directions. On the one hand, we will consider $\mathcal{EL}$ concepts w.r.t. a background ontology. On the other hand, we will consider a setting where the ABox contains not just concept assertions, but also role assertions. In the latter case, one can use not just generalization of

concepts, but also renaming of individuals as operations for achieving compliance (safety). Finally, of course, these two extensions should be combined.

References

[1] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.

[2] 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, New York, NY, USA, 2003.

[3] F. Baader, F. Kriegel, A. Nuradiansyah, and R. Peñaloza. Making repairs in description logics more gentle. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pages 319–328, 2018.

[4] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99)*, pages 96–101, 1999.

[5] F. Baader and B. Morawska. Unification in the description logic $\mathcal{EL}$. *Logical Methods in Computer Science*, 6(3), 2010.

[6] F. Baader and A. Nuradiansyah. Towards Privacy-Preserving Ontology Publishing. In M. Ortiz and T. Schneider, editors, *Proceedings of the 31st International Workshop on Description Logics (DL 2018)*, CEUR Workshop Proceedings, 2018.

[7] P. A. Bonatti. Fast compliance checking in an OWL2 fragment. In J. Lang, editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 1746–1752. ijcai.org, 2018.

[8] S. Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In R. L. de Mántaras and L. Saitta, editors, *Proc. of the 16th Eur. Conf. on Artificial Intelligence (ECAI 2004)*, pages 298–302, 2004.

[9] B. Cuenca Grau and E. V. Kostylev. Logical foundations of privacy-preserving publishing of linked data. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 943–949, 2016.

[10] T. Eiter and G. Gottlob. Hypergraph transversal computation and related problems in logic and AI. In S. Flesca, S. Greco, N. Leone, and G. Ianni, editors, *Logics in Artificial Intelligence, European Conference, JELIA 2002*, volume 2424 of *Lecture Notes in Computer Science*, pages 549–564. Springer, 2002.

[11] M. L. Fredman and L. Khachiyan. On the complexity of dualization of monotone disjunctive normal forms. *J. Algorithms*, 21(3):618–628, 1996.

[12] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4):14:1–14:53, 2010.

[13] G. Gottlob and E. Malizia. Achieving new upper bounds for the hypergraph duality problem through logic. *SIAM J. Comput.*, 47(2):456–492, 2018.

[14] I. Horrocks, L. Li, D. Turi, and S. Bechhofer. The instance store: DL reasoning with large numbers of individuals. In *Proceedings of the 2004 International Workshop on Description Logics (DL2004), Whistler, British Columbia, Canada, June 6-8, 2004*, 2004.

[15] F. Kriegel. The distributive, graded lattice of $\mathcal{EL}$ concept descriptions and its neighborhood relation (extended version). LTCS-Report 18-10, Chair of Automata Theory, Institute of Theoretical Computer Science, TU Dresden, Dresden, Germany, 2018. `https://tu-dresden.de/inf/lat/reports#Kr-LTCS-18-10`.

[16] R. Küsters. *Non-standard Inferences in Description Logics*, volume 2100 of *Lecture Notes in Artificial Intelligence*. Springer, 2001.