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

LTCS–Report

Computing Safe Anonymisations of Quantified ABoxes w.r.t. $\mathcal{EL}$ Policies (Extended Version)

Franz Baader, Francesco Kriegel,
Adrian Nuradiansyah, Rafael Peñaloza

LTCS-Report 20-09

This is an extended version of an article published in: Proceedings of the 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21), ACM

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

Computing Safe Anonymisations of Quantified ABoxes w.r.t. $\mathcal{EL}$ Policies (Extended Version)*

Franz Baader, Francesco Kriegel,
Adrian Nuradiansyah, Rafael Peñaloza

Abstract

In recent work, we have shown how to compute compliant anonymizations of quantified ABoxes w.r.t. $\mathcal{EL}$ policies. In this setting, quantified ABoxes can be used to publish information about individuals, some of which are anonymized. The policy is given by concepts of the Description Logic (DL) $\mathcal{EL}$, and compliance means that one cannot derive from the ABox that some non-anonymized individual is an instance of a policy concept. If one assumes that a possible attacker could have additional knowledge about some of the involved non-anonymized individuals, then compliance with a policy is not sufficient. One wants to ensure that the quantified ABox is safe in the sense that none of the secret instance information is revealed, even if the attacker has additional compliant knowledge. In the present paper, we show that safety can be decided in polynomial time, and that the unique optimal safe anonymization of a non-safe quantified ABox can be computed in exponential time, provided that the policy consists of a single $\mathcal{EL}$ concept.

1 Introduction

When making information about persons available online, one needs to ensure that certain privacy constraints described by a privacy policy are taken into account. The policy may be formulated by the data provider, the individuals whose data are to be published, or be due to some legal requirements. There is a large body of work on this topic in different areas of computer science [11], but here we restrict our attention to a setting where data about real-world individuals are to be published, but certain information about these individuals needs to be

*funded by the Deutsche Forschungsgemeinschaft (DFG) – Project number 430150274

kept secret. This differs from the setting of statistical databases (e.g., for medical research), where only anonymized and possibly aggregated data are published, but there is still the danger that information on real-world individuals can be extracted with a certain probability. Approaches for warding off this danger are, for example, k-anonymity [15] and differential privacy [10], but this is not what the current paper is about.

In the setting where the original data rather than statistical information about it are to be published, we further restrict our attention to work related to ontologies and RDF. There are two approaches for achieving privacy that have been investigated in that context. First, instead of making the data public, one can provide only restricted access through queries, whose answers are monitored by a "censor", which may decide not to give an answer or even lie if needed to satisfy the constraints [7, 6, 8]. Second, one can publish the data in an appropriately anonymized form, while keeping as much information about individuals as is allowed by the policy available [12, 13, 9, 2, 5, 3].

Here we follow the second approach. The works in this area differ from each other in several aspects. The papers [2, 5, 3] and this one allow for arbitrary modifications of the original data set, as long as the new data is logically implied by the original one. In contrast, the work from [12, 13, 9] restricts modifications to the application of certain anonymization operations. Another distinguishing criterion is which formalisms are employed for representing the data and the policy. While in the work described in [12, 13, 9] RDF graphs are used to represent the data and conjunctive queries to describe the policy, the papers [2, 5, 3] consider the setting where DL ABoxes represent the data and concepts of the DL $\mathcal{EL}$ describe the policy. More precisely, a restricted form of ABoxes, called instance store, is considered in [2, 5], whereas in [3] and in the present paper so-called quantified ABoxes are employed. Basically, quantified ABoxes extend traditional DL ABoxes by allowing for anonymized individuals, which from a logical point of view are represented as existentially quantified variables. Finally, one can distinguish approaches according to whether and which kind of attacker's knowledge is assumed to exist. Of the mentioned papers, only [3] does not allow for attacker's knowledge, i.e., restricts the attention purely to achieving compliance with the policy. Here, we employ the same formal setup as [3] but addresses safety. The only work where the formalisms for representing the attacker's knowledge and formalizing the data differ is [5].

Before diving into the technical details of our approach, let us illustrate the problem it solves by a simple example. Assume that Ben goes to a new school in fall, but does not want the people in the school to know that both of his parents are comedians. This privacy constraint can be formalized by the $\mathcal{EL}$ concept $P := \exists\, \mathit{mother}.\,(\mathit{Comedian} \sqcap \exists\, \mathit{spouse}.\, \mathit{Comedian})$. Ben needs to provide contact details of one parent, and decides to give his father's name since his mother never answers her mobile. This results in the quantified ABox

$$\exists\{x\}.\{\mathit{mother}(\mathit{BEN},x),\mathit{Comedian}(x),\mathit{spouse}(x,\mathit{JERRY}),\mathit{Comedian}(\mathit{JERRY})\},$$

where Ben's mother is represented by a variable since he did not disclose her name. Since this ABox is not compliant with Ben's policy P, he decides to hide the information that his father is a comedian. This yields the quantified ABox

$$\exists\{x\}.\{\mathit{mother}(\mathit{BEN},x),\mathit{Comedian}(x),\mathit{spouse}(x,\mathit{JERRY})\}, \tag{1}$$

which is compliant with P. However, this ABox is not safe for P since an attacker that knows $\mathit{Comedian}(\mathit{JERRY})$ (which on its own is compliant with P, and can easily be found out since Jerry is famous) can combine this knowledge with the given quantified ABox to derive that Ben is an instance of P. Had Ben instead removed the information that his (anonymized) mother is a comedian, and kept the information that Jerry is one, then the quantified ABox

$$\exists\{x\}.\{\mathit{mother}(\mathit{BEN},x),\mathit{spouse}(x,\mathit{JERRY}),\mathit{Comedian}(\mathit{JERRY})\} \tag{2}$$

obtained this way would again have been compliant with, but not safe for P. In fact, while an attacker could not obtain information about the anonymized individual x, and thus could not have learned $\mathit{Comedian}(x)$, other sources might have provided the information that Ben's mother is a comedian that is married to Jerry. The quantified ABox $\exists\{y\}.\{\mathit{mother}(\mathit{BEN},y),\mathit{Comedian}(y),\mathit{spouse}(y,\mathit{JERRY})\}$ representing this information is compliant, and adding it to the above ABox reveals that Ben is an instance of P. Thus, Ben needs to remove $\mathit{Comedian}(\mathit{JERRY})$ as well, which finally results in a quantified ABox that is safe for P:

$$\exists\{x\}.\{\mathit{mother}(\mathit{BEN},x),\mathit{spouse}(x,\mathit{JERRY})\}. \tag{3}$$

We show in this paper that, whether or not a given quantified ABox is safe for such a singleton policy, can be decided in polynomial time. In addition we describe how to compute an optimal safe generalization of a non-safe ABox in exponential time, where optimal means that the least amount of information is lost. In our example, the finally obtained safe ABox is actually not optimal.

2 Preliminaries

As mentioned earlier, a specific instance of the safety problem is determined by the available query language, which is used to formulate the safety policy, and the formalism for representing the data to be published. Following [3], we employ $\mathcal{EL}$ concepts as queries and represent the data as quantified ABoxes. The latter differ from the ABoxes usually employed in DL [1] in that (i) concept assertions are restricted to concept names, and (ii) existentially quantified variables can be used to represent anonymous individuals. While (ii) increases the expressive power of the formalism, (i) is not a real restriction since concept assertions involving complex concepts can be simulated based on the expressiveness provided by (ii).

More formally, we fix a signature Σ, which is the disjoint union of a set Σ_O of *object names*, a set Σ_C of *concept names*, and a set Σ_R of *role names*. A *quantified ABox* $\exists X.\mathcal{A}$ consists of a finite subset X of Σ_O and a *matrix* $\mathcal{A}$, which is a finite set containing *concept assertions* $A(u)$ and *role assertions* $r(u,v)$ where $u, v \in \Sigma_\mathsf{O}$, $A \in \Sigma_\mathsf{C}$, and $r \in \Sigma_\mathsf{R}$. The elements of X are called *variables*. An *individual name* in $\exists X.\mathcal{A}$ is an object name that occurs in $\mathcal{A}$ and is not a variable. We denote the set of these individual names as $\Sigma_\mathsf{I}(\exists X.\mathcal{A})$, or simply as Σ_I if the quantified ABox is clear from the context.[1] A traditional *ABox* is a quantified ABox where the quantifier prefix is empty. Instead of $\exists\emptyset.\mathcal{A}$ we simply write $\mathcal{A}$. The *matrix* $\mathcal{A}$ of a quantified ABox $\exists X.\mathcal{A}$ is such a traditional ABox.

The semantics of quantified ABoxes is defined using *interpretations*, which are of the form $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ (the *domain*) is a non-empty set and $\cdot^\mathcal{I}$ (the *interpretation function*) maps each object name u from Σ_O to an element $u^\mathcal{I}$ of $\Delta^\mathcal{I}$, each concept name A from Σ_C to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$, and each role name r from Σ_R to a binary relation $r^\mathcal{I}$ over $\Delta^\mathcal{I}$. The interpretation $\mathcal{I}$ is a *model* of $\exists X.\mathcal{A}$ if there is an interpretation $\mathcal{J}$ with the same the domain as $\mathcal{I}$ such that the interpretation functions $\cdot^\mathcal{J}$ and $\cdot^\mathcal{I}$ coincide on $\Sigma \setminus X$, $u^\mathcal{J} \in A^\mathcal{J}$ holds for each $A(u) \in \mathcal{A}$, and $(u^\mathcal{J}, v^\mathcal{J}) \in r^\mathcal{J}$ holds for each $r(u,v) \in \mathcal{A}$. The quantified ABox $\exists X.\mathcal{A}$ *entails* the quantified ABox $\exists Y.\mathcal{B}$ ($\exists X.\mathcal{A} \models \exists Y.\mathcal{B}$) if each model of $\exists X.\mathcal{A}$ is a model of $\exists Y.\mathcal{B}$.

Following [3], when considering two quantified ABoxes $\exists X.\mathcal{A}$ and $\exists Y.\mathcal{B}$, we henceforth assume without loss of generality that they are *renamed apart* in the sense that X is disjoint with $Y \cup \Sigma_\mathsf{I}(\exists Y.\mathcal{B})$ and Y is disjoint with $X \cup \Sigma_\mathsf{I}(\exists X.\mathcal{A})$.

As pointed out in [3], quantified ABoxes and conjunctive queries are essentially the same. In particular, ABox entailment coincides with query containment. It follows that the entailment problem for quantified ABoxes is NP-complete and that $\exists X.\mathcal{A}$ entails $\exists Y.\mathcal{B}$ iff there is a homomorphism from $\exists Y.\mathcal{B}$ to $\exists X.\mathcal{A}$. Such a *homomorphism* is a mapping $h\colon \Sigma_\mathsf{I}(\exists Y.\mathcal{B}) \cup Y \to \Sigma_\mathsf{I}(\exists X.\mathcal{A}) \cup X$ such that $h(a) = a$ for each $a \in \Sigma_\mathsf{I}(\exists Y.\mathcal{B})$,[2] and $A(u) \in \mathcal{B}$ implies $A(h(u)) \in \mathcal{A}$, and $r(u,v) \in \mathcal{B}$ implies $r(h(u), h(v)) \in \mathcal{A}$.

The set of $\mathcal{EL}$ *concept descriptions* over Σ is defined by induction: any concept name $A \in \Sigma_\mathsf{C}$ as well as $\top$ (top concept) belongs to this set, and if $r \in \Sigma_\mathsf{R}$ is a role name and C, D are known to belong to the set, then $C \sqcap D$ (conjunction) and $\exists r.C$ (existential restriction) belong to it as well. Given an interpretation $\mathcal{I}$, we extend $\cdot^\mathcal{I}$ to $\mathcal{EL}$ concept descriptions:

- $(\exists r.C)^\mathcal{I} := \{\, \delta \mid (\delta, \gamma) \in r^\mathcal{I} \text{ and } \gamma \in C^\mathcal{I} \text{ for some } \gamma \in \Delta^\mathcal{I} \,\}$;

[1] We use a, b, c for individual names, u, v, w for object names, and x, y, z for variables.

[2] More specifically, we require $h(a) = a$ and $a \in \Sigma_\mathsf{I}(\exists X.\mathcal{A})$ for each $a \in \Sigma_\mathsf{I}(\exists Y.\mathcal{B})$.

- $(C \sqcap D)^\mathcal{I} := C^\mathcal{I} \cap D^\mathcal{I}$.

Given $\mathcal{EL}$ concept descriptions C and D, we say that C is *subsumed* by D ($C \sqsubseteq_\emptyset D$) if $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds for each interpretation $\mathcal{I}$; C is *equivalent* to D ($C \equiv_\emptyset D$) if $C \sqsubseteq_\emptyset D$ and $D \sqsubseteq_\emptyset C$, and C is *strictly subsumed* by D ($C \sqsubset_\emptyset D$) if $C \sqsubseteq_\emptyset D$ and $C \not\equiv_\emptyset D$. The subscript $\emptyset$ in $\sqsubseteq_\emptyset$ indicates that no terminological axioms are available, i.e., we consider subsumption w.r.t. the empty TBox. If furthermore $\exists X.\mathcal{A}$ is a quantified ABox and u is an object name, then we say that u is an *instance* of C w.r.t. $\exists X.\mathcal{A}$ ($\exists X.\mathcal{A} \models C(u)$) if $u^\mathcal{I} \in C^\mathcal{I}$ is satisfied for each model $\mathcal{I}$ of $\exists X.\mathcal{A}$. The subsumption and the instance problem are known to be solvable in polynomial time [4, 3]. In particular, we have the following two recursive characterizations.

Lemma 1. *Let C, D be $\mathcal{EL}$ concept descriptions. Then $C \sqsubseteq_\emptyset D$ holds iff the following two statements are satisfied:*

1. *$A \in \mathsf{Conj}(D)$ implies $A \in \mathsf{Conj}(C)$ for each concept name A;*
2. *for each existential restriction $\exists r.F \in \mathsf{Conj}(D)$, there is an existential restriction $\exists r.E \in \mathsf{Conj}(C)$ such that $E \sqsubseteq_\emptyset F$.*

Lemma 2. *Let $\mathcal{A}$ be an ABox, C an $\mathcal{EL}$ concept description, and $u \in \Sigma_\mathsf{O}$. Then $\mathcal{A} \models C(u)$ holds iff the following two statements are satisfied:*

1. *for each concept name $A \in \mathsf{Conj}(C)$, the ABox $\mathcal{A}$ contains $A(u)$;*
2. *for each existential restriction $\exists r.D \in \mathsf{Conj}(C)$, the ABox $\mathcal{A}$ contains a role assertion $r(u, v)$ such that $\mathcal{A} \models D(v)$.*

An $\mathcal{EL}$ *atom* is either a concept name A or an existential restriction $\exists r.C$. Clearly, any $\mathcal{EL}$ concept description C is a conjunction of atoms. We call this the *top-level conjunction* of C and denote the set of atoms occurring it as $\mathsf{Conj}(C)$. The set of atoms occurring as subconcepts of C is defined as $\mathsf{Atoms}(C) := \mathsf{Conj}(C) \cup \bigcup\{\, \mathsf{Atoms}(D) \mid \exists r.D \in \mathsf{Conj}(C) \,\}$. We will also employ the reduced forms C^r of $\mathcal{EL}$ concept descriptions C [14], which are defined as follows: $A^r := A$ for $A \in \Sigma_\mathsf{C}$; $(\exists r.C)^r := \exists r.C^r$; and $(C \sqcap D)^r := C^r$ if $C \sqsubseteq_\emptyset D$, $(C \sqcap D)^r := D^r$ if $D \sqsubset_\emptyset C$, and $(C \sqcap D)^r := C^r \sqcap D^r$ if C and D are incomparable w.r.t. subsumption. As shown in [14], $C \equiv_\emptyset C^r$ and $C \equiv_\emptyset D$ implies that C^r and D^r are equal up to associativity and commutativity of conjunction.

Finally, let us come back to the claim that concept assertions $C(a)$ involving complex concept descriptions C can be expressed by quantified ABoxes. To that purpose, we view $\mathcal{EL}$ concept descriptions as trees and use paths in these trees as variables. More formally, a *path* in an $\mathcal{EL}$ concept description C is a sequence $p = D_0 \xrightarrow{r_1} D_1 \xrightarrow{r_2} \ldots \xrightarrow{r_n} D_n$ such that $D_0 = C$ and $\exists r_i.D_i \in \mathsf{Conj}(D_{i-1})$ for each index $i \in \{1, \ldots, n\}$. We call $\mathsf{target}(p) := D_n$ the *target* of p. Note that $n = 0$ is possible, i.e., C is always a path in C, called the *root*. The set of all paths in C is denoted by $\mathsf{Paths}(C)$. By viewing the elements of $\mathsf{Paths}(C) \setminus \{C\}$ as new object names, the *ABox translation* of $C(a)$ can be defined as the quantified ABox $\exists(\mathsf{Paths}(C) \setminus \{C\}).\mathcal{A}_{C(a)}$ where, for all paths $p, q \in \mathsf{Paths}(C)$, $A(p)$ is in $\mathcal{A}_{C(a)}$ if $A \in \mathsf{Conj}(\mathsf{target}(p))$ and where $r(p,q)$ is in $\mathcal{A}_{C(a)}$ if q extends p with one r-edge, i.e., if $q = p \xrightarrow{r} D$ for some $\exists r.D \in \mathsf{Conj}(\mathsf{target}(p))$, and where we finally replace each occurrence of C in position of an object name in $\mathcal{A}_{C(a)}$ with the individual name a. Note that this quantified ABox contains a as the only individual name, whereas all paths in $\mathsf{Paths}(C) \setminus \{C\}$ are variables. It is clearly equivalent to the assertion $C(a)$.

Similarly, we define the *ABox translation* of an $\mathcal{EL}$ concept description C as the quantified ABox $\exists\mathsf{Paths}(C).\mathcal{A}_C$ where $A(p)$ is in $\mathcal{A}_C$ if $A \in \mathsf{Conj}(\mathsf{target}(p))$ and where $r(p,q)$ is in $\mathcal{A}_C$ if $q = p \xrightarrow{r} D$ for some $\exists r.D \in \mathsf{Conj}(\mathsf{target}(p))$. Note that, in the case where C is a concept name A, the symbol A would need to be both a concept name and an object name, which is forbidden by the very definition of a signature. Without loss of generality we assume that in such a case the occurrences of the symbol A in place of an object name is suitably replaced by another symbol.

3 A Characterization of Safety

We define the notions of compliance and safety, and then give a characterization of safety for the case of singleton policies. This characterization provides us with a polynomial time decision procedure for safety in this restricted setting. The exact complexity of deciding safety in the general case is still open, though it is easy to show an NP upper bound using ideas from [12, 13].

A *policy* $\mathcal{P}$ is a finite set of $\mathcal{EL}$ concept descriptions. A quantified ABox $\exists X.\mathcal{A}$ is *compliant* with $\mathcal{P}$ if it does not contain an individual name that belongs to a concept in $\mathcal{P}$, i.e., there does not exist a policy concept $P \in \mathcal{P}$ and an individual name $a \in \Sigma_{\mathsf{I}}(\exists X.\mathcal{A})$ such that $\exists X.\mathcal{A} \models P(a)$. Testing for compliance thus boils down to solving the instance problem, and can consequently be realized in polynomial time.

Safety is a stronger notion, which requires compliance to be preserved under addition of any compliant data. More formally, $\exists X.\mathcal{A}$ is *safe* for the policy $\mathcal{P}$ if, for each quantified ABox $\exists Y.\mathcal{B}$ that is compliant with $\mathcal{P}$ and renamed apart from $\exists X.\mathcal{A}$, the union $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B} := \exists(X \cup Y).(\mathcal{A} \cup \mathcal{B})$ is compliant with $\mathcal{P}$. Since the empty ABox is always compliant and renamed apart, safety for $\mathcal{P}$ implies compliance with $\mathcal{P}$, but the opposite implication need not hold, as illustrated by our example in the introduction.

The goal of this section is to find necessary and sufficient conditions for safety in the case where the policy is a *singleton* set, i.e., $\mathcal{P} = \{P\}$ for an $\mathcal{EL}$ concept description P, where we assume w.l.o.g. that P is reduced. We also assume that P is not $\top$ and that the given quantified ABox $\exists X. \mathcal{A}$ contains at least one individual name since otherwise safety is trivial to decide.

In [2], safety was investigated for data represented by $\mathcal{EL}$ instance stores, i.e., by traditional ABoxes with complex concept assertions, but without role assertions. The results proved in [2] can be used to derive the following characterization of safety for general policies: a given instance store is safe for the policy $\mathcal{P}$ iff it is compliant with $\mathsf{Conj}(\mathcal{P}) := \bigcup\{\, \mathsf{Conj}(P) \mid P \in \mathcal{P} \,\}$. This characterization reduces safety in polynomial time to compliance.

In our setting, compliance with the top-level conjuncts of the policy concept is still a necessary condition for safety, but it is no longer sufficient. In fact, it is easy to see that each quantified ABox that is safe for $\{P\}$ must also be compliant with $\mathsf{Conj}(P)$. Assume that C is a top-level conjunct of the policy concept P such that $\exists X. \mathcal{A}$ entails $C(a)$. We write $P \setminus C$ for the concept obtained from P by deleting C from the top-level conjunction. Now assume that $\exists Y. \mathcal{B}$ is the ABox translation of $(P \setminus C)(a)$. Since the policy concept P is assumed to be reduced, we infer that $(P \setminus C) \not\sqsubseteq_\emptyset C$, which implies that $\exists Y. \mathcal{B}$ is compliant with $\{P\}$. However, the union of $\exists X. \mathcal{A}$ and $\exists Y. \mathcal{B}$ clearly entails $P(a)$.

Example 3. To illustrate the above observation, we consider the policy concept $P := A \sqcap B \sqcap \exists r. A$. The ABox $\exists \emptyset. \{A(a)\}$ is compliant with $\{P\}$, but it entails $A(a)$ for the top-level conjunct A of P. This ABox is not safe for $\{P\}$ since, on the one hand, the ABox $\exists \{x\}. \{B(a),\, r(a, x),\, A(x)\}$ complies with $\{P\}$, but, on the other hand, its union with $\exists \emptyset. \{A(a)\}$ entails that a is an instance of P. Note that the second ABox $\exists \{x\}. \{B(a),\, r(a, x),\, A(x)\}$ is (equivalent to) the ABox translation of $(P \setminus A)(a) = (B \sqcap \exists r. A)(a)$.

Due to the presence of role assertions, safety enforces an even stronger condition. Not only the atoms appearing in the top-level conjunction of P need to be considered, but all atoms occurring somewhere in P, i.e., all elements of $\mathsf{Atoms}(P)$. Such an atom is either a concept name or an existential restriction.

First, consider a concept name A that occurs in the policy concept P, i.e., $A \in \mathsf{Atoms}(P)$. The case where A is a top-level conjunct has already been investigated above. So assume that A is not in the top-level conjunction of P, i.e., there is a path p in P with at least one edge such that A is in $\mathsf{Conj}(\mathsf{target}(p))$, and assume that $\exists X. \mathcal{A}$ entails $A(a)$. Construct the ABox $\exists Y. \mathcal{B}$ by taking the ABox translation of $P(b)$, for a fresh individual name b, but removing the concept assertion $A(p)$ and then replacing p with a. The remaining information on a in $\exists Y. \mathcal{B}$, which is the concept $\mathsf{target}(p) \setminus A$, cannot be subsumed by the policy concept description P (since the role depth[3] of $\mathsf{target}(p)$ is strictly smaller than the role depth of P). Furthermore, b cannot be an instance of P (since P is reduced and we have removed one occurrence of A). It follows that $\exists Y. \mathcal{B}$ is

[3] The role depth of an $\mathcal{EL}$ concept description is the maximal nesting of existential restrictions in this description.

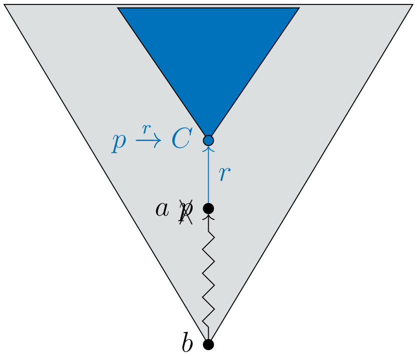


Figure 1: Constructing a counterexample against safety when the ABox does not comply with an atom $\exists r.\, C$ occurring in the policy concept description.

compliant with $\{P\}$, but its union with $\exists X.\, \mathcal{A}$ is not since it reveals the sensitive information that b is an instance of P.

Example 4. Consider the policy concept $P \coloneqq B \sqcap \exists r.\, A$, for which the concept name A is an element of $\mathsf{Atoms}(P)$. In particular, A is a top-level conjunct of the target of the path $P \xrightarrow{r} A$. The ABox $\exists \emptyset.\, \{A(a)\}$ entails $A(a)$ and it is not safe for $\{P\}$. To see the latter, note that the ABox $\exists \emptyset.\, \{B(b),\ r(b,a)\}$ is compliant with $\{P\}$, and that its union with $\exists \emptyset.\, \{A(a)\}$ entails $P(b)$. The second ABox $\exists \emptyset.\, \{B(b),\ r(b,a)\}$ was exactly obtained by applying the general construction sketched above to this specific example.

For an existential restriction $\exists r.\, C$ instead of the concept name A, we proceed in a similar way, except that during the construction of $\exists Y.\, \mathcal{B}$ we do not remove $A(p)$, but instead remove the assertion $r(p, p \xrightarrow{r} C)$ as well as all assertions involving a path with prefix $p \xrightarrow{r} C$. This corresponds to removing from the ABox translation the part corresponding to the subconcept C. This construction is depicted in Figure 1, where the gray area depicts the parts remaining in the counterexample ABox $\exists Y.\, \mathcal{B}$, while the blue area is removed.

Example 5. Take $P \coloneqq B \sqcap \exists s.\, \exists r.\, \top$ as the policy concept. $\mathsf{Atoms}(P)$ contains the existential restriction $\exists r.\, \top$. More specifically, $\exists r.\, \top$ is in $\mathsf{Conj}(\mathsf{target}(P \xrightarrow{s} \exists r.\, \top))$. The quantified ABox $\exists \{x\}.\, \{r(a,x)\}$ entails $\exists r.\, \top(a)$. The construction sketched above yields the ABox $\exists \emptyset.\, \{B(b),\ s(b,a)\}$. This ABox clearly complies with $\{P\}$, but its union with $\exists \{x\}.\, \{r(a,x)\}$ entails $P(b)$. Consequently, $\exists \{x\}.\, \{r(a,x)\}$ is not safe for $\{P\}$.

Summing up, we have seen that safety for $\{P\}$ implies compliance with the extended policy $\mathsf{Atoms}(P)$, which contains each atom C that is a top-level conjunct of $\mathsf{target}(p)$ for some path p in the policy concept P. Distinguishing between the two types of atoms and using the characterization of the instance problem given by Lemma 6 in [3], this fact can be stated as follows.

Lemma 6. *If $\exists X.\mathcal{A}$ is safe for $\{P\}$, then $\exists X.\mathcal{A}$ is compliant with $\mathsf{Atoms}(P)$, i.e., the following two conditions are satisfied:*

1. *For each individual name a and for each concept name $A \in \mathsf{Atoms}(P)$, the concept assertion $A(a)$ is not in $\mathcal{A}$.*
2. *For each individual name a, for each role assertion $r(a, u)$ in $\mathcal{A}$, and for each existential restriction $\exists r.C$ in $\mathsf{Atoms}(P)$, the matrix $\mathcal{A}$ does not entail $C(u)$.*

It turns out, however, that compliance with $\mathsf{Atoms}(P)$ is still not sufficient to ensure safety for $\{P\}$. A counterexample is the following, which illustrates that it is not necessary to find a whole element of $\mathsf{Atoms}(P)$ in the ABox to lose safety.

Example 7. Consider $\exists X.\mathcal{A} \coloneqq \exists\{x\}.\{r(a, x),\ A(x),\ s(x, b)\}$ and the policy concept $P \coloneqq A \sqcap \exists r.(A \sqcap \exists s.A)$. Note that $\exists X.\mathcal{A}$ is compliant with $\mathsf{Atoms}(P) = \{\exists r.(A \sqcap \exists s.A),\ \exists s.A,\ A\}$. However, $\exists X.\mathcal{A}$ is not safe for $\{P\}$: for the ABox $\exists Y.\mathcal{B} \coloneqq \exists\emptyset.\{A(a),\ A(b)\}$, which is compliant with $\{P\}$, the union $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B}$ entails $P(a)$. The reason is that, while we do not find the whole atom $\exists r.(A \sqcap \exists s.A)$ in $\exists X.\mathcal{A}$, we find the part $\exists r.(A \sqcap \exists s.\top)$. The concept name A missing in the existential restriction $\exists s.\top$ is added by the assertion $A(b)$ in the attacker ABox $\exists Y.\mathcal{B}$.

To formalize what it means to "find part of an atom" in a quantified ABox, we will use the notion of a partial homomorphism. To motivate this notion, we first reformulate the second condition in Lemma 6 using the following homomorphism characterization of the instance problem, which is an easy consequence of Lemma 6 in [3]. For a quantified ABox $\exists X.\mathcal{A}$, the matrix $\mathcal{A}$ entails $C(u)$ iff there is a homomorphism from C to $\exists X.\mathcal{A}$ at u, which is a mapping $h\colon \mathsf{Paths}(C) \to \Sigma_\mathsf{I} \cup X$ satisfying the following conditions:

1. $h(C) = u$
2. For each $p \in \mathsf{Paths}(C)$, the following two conditions hold:
   (a) $A(j(p)) \in \mathcal{A}$ for each concept name $A \in \mathsf{Conj}(\mathsf{target}(p))$,
   (b) $r(j(p), j(p \xrightarrow{r} D)) \in \mathcal{A}$ for each existential restriction $\exists r.D \in \mathsf{Conj}(\mathsf{target}(p))$.

The second condition in Lemma 6 can now be reformulated as

2. For each individual name a, for each role assertion $r(a, u)$ in $\mathcal{A}$, and for each existential restriction $\exists r.C$ in $\mathsf{Atoms}(P)$, there is no homomorphism from C to $\exists X.\mathcal{A}$ at u.

The idea is now to replace "homomorphism" in this condition with "partial homomorphism." Intuitively, a partial homomorphism is *almost* a homomorphism from the concept C to the quantified ABox $\exists X.\mathcal{A}$ at u, which can, however, omit mapping some parts of C into the ABox in case the ABox has an individual at

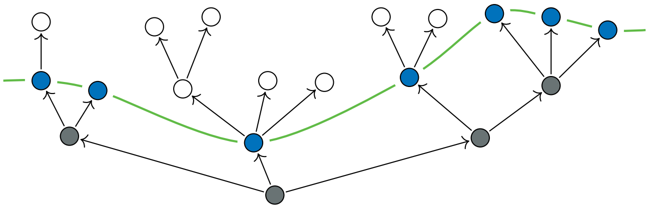


Figure 2: An ideal induced by a maximal antichain in a tree

the “cut-off points.” In order to give a more formal definition of partial homomorphisms, we first need to introduce some auxiliary notions. The set $\mathsf{Paths}(C)$ of all paths in an $\mathcal{EL}$ concept description C is partially ordered by the prefix relation $\leq$. The smallest path is C (the root) and the maximal paths are those $p \in \mathsf{Paths}(C)$ where $\mathsf{Conj}(\mathsf{target}(p))$ does not contain any existential restriction, which we call *leaves*. Each subset $\mathfrak{X} \subseteq \mathsf{Paths}(C)$ induces an *ideal* $\downarrow\mathfrak{X} := \{\, p \mid p \leq q$ for some $q \in \mathfrak{X} \,\}$. Furthermore, an *antichain* is a subset $\mathfrak{A} \subseteq \mathsf{Paths}(C)$ such that no two paths in $\mathfrak{A}$ are comparable w.r.t. $\leq$. An antichain is *maximal* if there is no strict superset that is an antichain as well. A maximal antichain $\mathfrak{A}$ corresponds to a cut through the syntax tree of C. Figure 2 gives an abstract visualization of a maximal antichain and the ideal induced by it: the antichain consists of the blue nodes and its induced ideal consists of all non-white nodes. The white nodes are pruned away by the cut.

Definition 8. Let C be an $\mathcal{EL}$ concept description and $\exists X.\mathcal{A}$ be a quantified ABox in which u is an object. A *partial homomorphism* from C to $\exists X.\mathcal{A}$ at u is a pair $(j, \mathfrak{B})$ consisting of a maximal antichain $\mathfrak{B}$ of $(\mathsf{Paths}(C), \leq)$, called the *border*, and a mapping $j\colon \downarrow\mathfrak{B} \to \Sigma_\mathsf{I} \cup X$ such that the following conditions are satisfied.

1. $j(C) = u$
2. If $p \in \downarrow\mathfrak{B} \setminus \mathfrak{B}$ (i.e., p is strictly below the border), then $j(p) \in X$.
3. If $p \in \mathfrak{B} \setminus \mathsf{Max}_\leq(\mathsf{Paths}(C))$ (i.e., p is in the border but is not a leaf), then $j(p) \in \Sigma_\mathsf{I}$.
4. If $p \in \downarrow\mathfrak{B}$ (i.e., p is in the border or is below the border) and $j(p) \in X$, then the following two conditions are satisfied:
   (a) $A(j(p)) \in \mathcal{A}$ for each concept name $A \in \mathsf{Conj}(\mathsf{target}(p))$,
   (b) $r(j(p), j(p \xrightarrow{r} D)) \in \mathcal{A}$ for each existential restriction $\exists r.D \in \mathsf{Conj}(\mathsf{target}(p))$.

Intuitively, a partial homomorphism only maps paths in C between the root and the border to objects of the ABox $\exists X.\mathcal{A}$. Cut-off points (paths $p \in \mathfrak{B} \setminus \mathsf{Max}_\leq(\mathsf{Paths}(C)))$ are mapped to individuals.

Example 9. Consider the concept description

$$C := A \sqcap \exists r.(B \sqcap \exists s.(\exists r.A \sqcap \exists r.B)) \sqcap \exists s.(A \sqcap \exists r.A \sqcap \exists r.B),$$

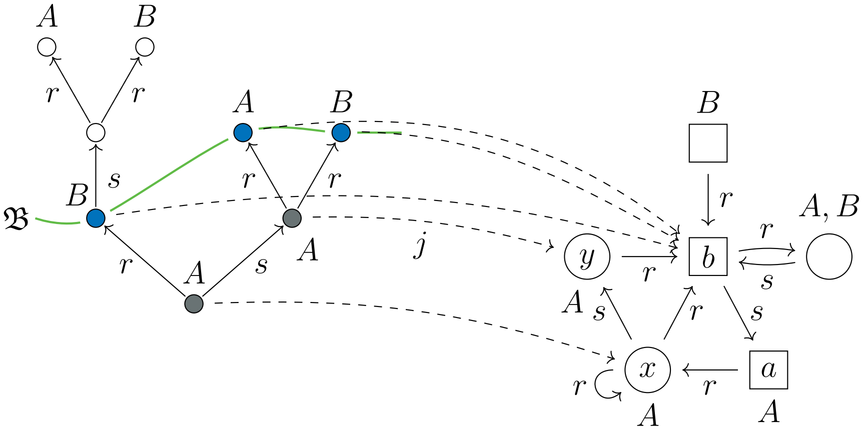


Figure 3: A partial homomorphism $(j, \mathfrak{B})$

which is depicted on the left-hand side of Figure 3. The three blue nodes form a maximal antichain, where for instance the right-most blue node represents the path $C \xrightarrow{s} A \sqcap \exists r.\, A \sqcap \exists r.\, B \xrightarrow{r} B$. Denote this antichain by $\mathfrak{B}$. The induced ideal $\downarrow\mathfrak{B}$ contains all non-white nodes. Consider now the ABox $\exists X.\, \mathcal{A}$ shown on the right-hand side of Figure 3, which contains the assertions $r(a, x)$, $A(x)$, among others. The pair $(j, \mathfrak{B})$ is a partial homomorphism from C to $\exists X.\, \mathcal{A}$ at x, where the mapping j is represented by the dashed lines in Figure 3.

Returning to Example 7, we see that the filler $A \sqcap \exists s.\, A$ of the existential restriction $\exists r.\, (A \sqcap \exists s.\, A) \in \mathsf{Atoms}(P)$ can be partially homomorphically mapped to the ABox $\exists X.\, \mathcal{A}$ at x via the partial homomorphism $(j, \mathfrak{B})$ where $\mathfrak{B} = \{A \sqcap \exists s.\, A \xrightarrow{s} A\}$ and j is defined by setting $j(A \sqcap \exists s.\, A) \coloneqq x$ and $j(A \sqcap \exists s.\, A \xrightarrow{s} A) \coloneqq b$. Moreover, $\mathcal{A}$ contains the role assertion $r(a, x)$ where a is an individual. This role assertion together with the partial homomorphism can be used to construct a compliant quantified ABox $\exists Y.\, \mathcal{B}$ that successfully attacks $\exists X.\, \mathcal{A}$. In fact, it suffices to know the remaining parts of the policy concept $A \sqcap \exists r.\, (A \sqcap \exists s.\, A)$ that are not homomorphically mapped to $\exists X.\, \mathcal{A}$, which is the top-level conjunct A and the concept name A within the existential restriction $\exists s.\, A$. These two parts are put into $\mathcal{B}$ through the assertions $A(a)$ and $A(b)$. As pointed out in Example 7, the quantified ABox $\exists Y.\, \mathcal{B}$ obtained this way complies with $\{P\}$, but its union with $\exists X.\, \mathcal{A}$ is no longer compliant.

We will show that the construction of an attacking quantified ABox is possible not only in this concrete example, but in general whenever such a situation occurs. To be more precise, assume that there is some existential restriction $\exists r.\, C \in \mathsf{Atoms}(P)$ (which is a top-level conjunct of $\mathsf{target}(p)$ for some path p in the policy concept P) and some role assertion $r(a, u) \in \mathcal{A}$ for an individual a such that there exists a partial homomorphism $(j, \mathfrak{B})$ from C to $\exists X.\, \mathcal{A}$ at u. Then it is possible to construct an attacking quantified ABox in a way similar to the one depicted in Figure 1. The only difference is that we do not cut out the whole

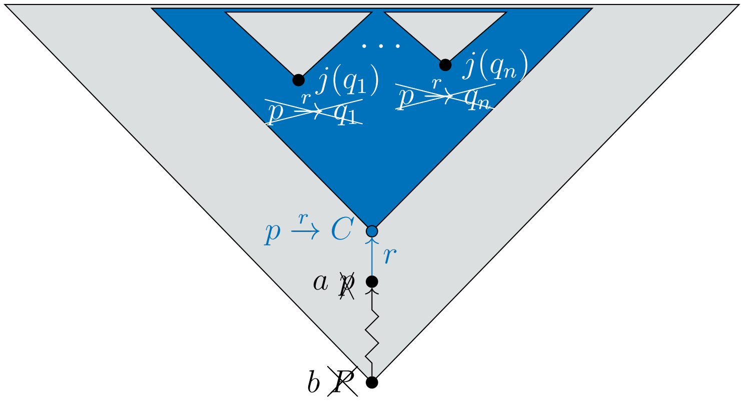


Figure 4: Constructing a counterexample against safety for the case where the ABox allows for a partial homomorphism

concept C but only those parts that are already present in the ABox due to the partial homomorphism. This idea is depicted in Figure 4. Thus, we strengthen the second condition in Lemma 6 as follows; in particular, the below lemma supersedes Lemma 6.

Lemma 10. *If $\exists X.\mathcal{A}$ is safe for $\{P\}$, then the following conditions are satisfied:*

1. *For each individual name a and for each concept name $A \in \mathsf{Atoms}(P)$, the concept assertion $A(a)$ is not in $\mathcal{A}$.*
2. *For each individual name a, for each role assertion $r(a,u)$ in $\mathcal{A}$, and for each existential restriction $\exists r.C$ in $\mathsf{Atoms}(P)$, there is no partial homomorphism from C to $\exists X.\mathcal{A}$ at u.*

Proof. We prove the contraposition and we start with the more involved second condition. Assume that there is some individual name a, some role assertion $r(a,u)$ in $\mathcal{A}$, and an existential restriction $\exists r.C$ in $\mathsf{Atoms}(P)$ such that there is a partial homomorphism $(j,\mathfrak{B})$ from C to $\exists X.\mathcal{A}$ at u. Since $\exists r.C$ is in $\mathsf{Atoms}(P)$, there exists a path $p \in \mathsf{Paths}(P)$ such that $\exists r.C \in \mathsf{Conj}(\mathsf{target}(p))$.

We have briefly described above how an ABox $\exists Y.\mathcal{B}$ can be constructed from the partial homomorphism. Formally, the construction is as follows.

(I) Initialize $\exists Y.\mathcal{B}$ as the ABox translation $\exists\mathsf{Paths}(P).\mathcal{A}_P$, cf. Page 6.

(II) Recall that the role assertion $r(a,u)$ is in $\mathcal{A}$. So we remove the corresponding role assertion $r(p, p \xrightarrow{r} C)$ from $\exists Y.\mathcal{B}$.

Since $p \xrightarrow{r} C$ is a path in P and $j^{-1}(X)$[4] is a subset of $\mathsf{Paths}(C)$, also $p \xrightarrow{r} q$ is a path in P for each path $q \in j^{-1}(X)$. We now remove from $\exists Y.\mathcal{B}$ each

[4] Recall that, for a mapping $f\colon P \to Q$ between two sets P and Q, the *pre-image* is defined as the mapping $f^{-1}\colon \wp(Q) \to \wp(P)$, $X \mapsto \{\, p \mid f(p) \in X \,\}$.

axiom involving a path $p \xrightarrow{r} q$ for some $q \in j^{-1}(X)$, i.e., we remove the following subset:

$$\{\, A(p \xrightarrow{r} q) \mid A \in \mathsf{Conj}(\mathsf{target}(q)) \text{ for some } q \in j^{-1}(X) \,\}$$
$$\cup \{\, r(p \xrightarrow{r} q, p \xrightarrow{r} q \xrightarrow{s} D) \mid \exists s.\, D \in \mathsf{Conj}(\mathsf{target}(q)) \text{ for some } q \in j^{-1}(X) \,\}.$$

Afterwards, we also remove all $p \xrightarrow{r} q$ for $q \in j^{-1}(X)$ from the variable set Y.

Note that the partial homomorphism $(j, \mathfrak{B})$ induces the following subset of $\exists X.\mathcal{A}$, which corresponds to the above removed subset:

$$\{\, A(j(q)) \mid A \in \mathsf{Conj}(\mathsf{target}(q)) \text{ for some } q \in j^{-1}(X) \,\}$$
$$\cup \{\, r(j(q), j(q \xrightarrow{s} D)) \mid \exists s.\, D \in \mathsf{Conj}(\mathsf{target}(q)) \text{ for some } q \in j^{-1}(X) \,\}.$$

(III) Let $q_1, \ldots, q_n$ be those paths from $\mathfrak{B}$ where $j(q_i) \in \Sigma_{\mathsf{I}}$. Then remove each $p \xrightarrow{r} q_i$ from Y and replace in $\mathcal{B}$ each $p \xrightarrow{r} q_i$ with the individual name $j(q_i)$.

Replace the variable p with a and remove p from Y. If $p \neq P$, then choose some fresh individual name b such that $b \neq a$ and $b \neq j(q_i)$ for each i, replace the variable P by b, and remove P from Y.

This construction is depicted in Figure 4; the gray area indicates which part remains in the counterexample ABox $\exists Y.\mathcal{B}$ while the blue area is removed.

First note that $\exists Y.\mathcal{B}$ is always acyclic—more specifically, it is a forest with roots $j(q_1)$, …, $j(q_n)$, and b (if $p \neq P$) or a (if $p = P$). We need to show that $\exists Y.\mathcal{B}$ is compliant with $\{P\}$. For each individual name $c \in \Sigma_{\mathsf{I}} \setminus \{a, b\}$, the most specific concept of which c is an instance is $\bigsqcap\{\, \mathsf{target}(q) \mid j(q) = c \,\}$. Thus, c is an instance of P if and only if $\bigsqcap\{\, \mathsf{target}(q) \mid j(q) = c \,\}$ is subsumed by P. However, the latter subsumption can never be satisfied, simply because the role depth of each $\mathsf{target}(q)$ is strictly smaller than the role depth of P.

Analogously, $(\mathsf{target}(p) \setminus \exists r.C) \sqcap \bigsqcap\{\, \mathsf{target}(q) \mid j(q) = a \,\}$ is the most specific concept of which a is an instance. Since P is reduced, we infer that $\mathsf{target}(p) \setminus \exists r.C \not\sqsubseteq_\emptyset \exists r.C$. Furthermore, we have that $\bigsqcap\{\, \mathsf{target}(q) \mid j(q) = a \,\} \not\sqsubseteq_\emptyset \exists r.C$, since each q is a path in C, i.e., the role depth of each $\mathsf{target}(q)$ is strictly smaller than $\mathsf{rd}(\exists r.C)$. We conclude that a is no instance of $\exists r.C$.

If $p \neq P$, then the above most specific concept of a cannot be subsumed by P since its role depth is too small, i.e., a is no instance of P. If $p = P$ (i.e., $\mathsf{target}(p) = P$), then a cannot be an instance of P simply because $\exists r.C$ is a top-level conjunct of P of which a is already not an instance.

It remains to show that, in the case $p \neq P$, the individual b is no instance of the policy concept P. Assume to the contrary that b is an instance of P, i.e., there exists a homomorphism h from (the ABox translation of) P to $\exists Y.\mathcal{B}$ at b. We immediately conclude that $h(P) = b$, and we will show in the following that

$h(p) = a$. For this purpose we first show by induction that $h(p') = p'$ holds true for each strict prefix p' of p except P (i.e., where $P < p' < p$). W.l.o.g. let

$$p = P \xrightarrow{r_1} C_1 \xrightarrow{r_2} \dots \xrightarrow{r_n} C_n.$$

We define $p_i \coloneqq P \xrightarrow{r_1} C_1 \xrightarrow{r_2} \dots \xrightarrow{r_i} C_i$ for each $i \in \{0, \dots, n\}$. The p_i for $i < n-1$ are exactly the strict prefixes of p (where p_0 equals P) and p_n equals p.

If we had $h(p_i \xrightarrow{r_{i+1}} C_{i+1}) = p_i \xrightarrow{r_{i+1}} D_{i+1}$ for some $\exists r_{i+1}. D_{i+1} \in \mathsf{Conj}(\mathsf{target}(p_i)) \setminus \{\exists r_{i+1}. C_{i+1}\}$, then the restriction of h to paths with prefix $p_i \xrightarrow{r_{i+1}} C_{i+1}$ would essentially be a homomorphism from C_{i+1} to D_{i+1}, which would certify that D_{i+1} is subsumed by C_{i+1}—a contradiction, since the policy concept P is reduced. In particular, note that during the construction of $\exists Y. \mathcal{B}$ we have not modified any axiom involving a path with prefix $p_i \xrightarrow{r_{i+1}} D_{i+1}$, i.e., the (maximal) sub-ABox of $\exists Y. \mathcal{B}$ containing only objects that are paths with prefix $p_i \xrightarrow{r_{i+1}} D_{i+1}$ is isomorphic to (the ABox translation of) D_{i+1}. By induction, we infer that $h(p_i) = p_i$ for each index $i \in \{1, \dots, n-1\}$ and further that $h(p) = a$.

We conclude that b can only be an instance of P if a is an instance of $\exists r. C$, which is not the case as we have already shown above.

By construction, the union of $\exists X. \mathcal{A}$ and $\exists Y. \mathcal{B}$ entails $P(b)$ if $P \neq p$ and entails $P(a)$ otherwise; a homomorphism h from P to $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B}$ at b or at a, respectively, is as follows:

$$\begin{aligned}
h(P) &\coloneqq \begin{cases} b & \text{if } P \neq p \\ a & \text{otherwise} \end{cases} \\
h(p) &\coloneqq a \\
h(p') &\coloneqq p' \text{ for each } p' \in \mathsf{Paths}(P) \setminus \{P, p\} \text{ with } p \xrightarrow{r} C \not\leq p' \\
h(p \xrightarrow{r} q) &\coloneqq j(q) \text{ for each } q \in \downarrow \mathfrak{B} \\
h(p \xrightarrow{r} q') &\coloneqq p \xrightarrow{r} q' \text{ for each } q' \text{ with } q < q' \text{ for some } q \in \mathfrak{B}
\end{aligned}$$

Thus, the ABox $\exists X. \mathcal{A}$ is not safe for $\{P\}$. Note that the homomorphism h can be seen as an extension of the mapping j of the partial homomorphism $(j, \mathfrak{B})$—it maps each path $p \xrightarrow{r} q$ where $q \in \downarrow \mathfrak{B}$ into the ABox $\exists X. \mathcal{A}$ (using j), and maps the remaining paths into $\exists Y. \mathcal{B}$. To see that h is indeed a homomorphism, recall that we have constructed $\exists Y. \mathcal{B}$ from the ABox translation of P by cutting out the part of $\exists r. C$ which is already homomorphically mapped to $\exists X. \mathcal{A}$ by means of the partial homomorphism. Leaving individual names aside, h is just the identical mapping into $\exists Y. \mathcal{B}$ with the exception that the part of $\exists r. C$ is mapped via j into $\exists X. \mathcal{A}$. The identical portion of h clearly satisfies the conditions of a homomorphism, cf. Page 9, and the j-portion of h satisfies those conditions since j already does so, cf. Definition 8.

For the remaining case, let a be an individual name and A a concept name in $\mathsf{Atoms}(P)$ such that the matrix $\mathcal{A}$ contains the concept assertion $A(a)$. Since A is in $\mathsf{Atoms}(P)$, there exists a path $p \in \mathsf{Paths}(P)$ such that $A \in \mathsf{Conj}(\mathsf{target}(p))$. We construct an ABox $\exists Y.\mathcal{B}$ similarly as above, but we replace Steps (II) and (III) with the following instructions:

(I) Initialize $\exists Y.\mathcal{B}$ as the ABox translation $\exists \mathsf{Paths}(P).\mathcal{A}_P$, cf. Page 6.

(II) Remove the concept assertion $A(p)$ from $\mathcal{B}$.

(III) Replace the variable p with a and remove p from Y. If $p \neq P$, then choose some fresh individual name b such that $b \neq a$, replace the variable P by b, and remove P from Y.

Similarly as above, we conclude that $\exists Y.\mathcal{B}$ is a counterexample against safety of $\exists X.\mathcal{A}$.

First of all, $\exists Y.\mathcal{B}$ is acyclic; more specifically, it is a tree with root b (if $p \neq P$) or root a (if $p = P$). The most specific concept of which a is an instance is $\mathsf{target}(p) \setminus A$. In the case $p \neq P$, this concept cannot be subsumed by P simply because its role depth is too small, which yields that a is no instance of P. If $p = P$ (i.e., $\mathsf{target}(p) = P$), then the root a cannot be an instance of A since $P \setminus A \not\sqsubseteq_\emptyset A$, which yields that a cannot be an instance of P since A is a top-level conjunct of P.

It remains to show that, in the case $p \neq P$, the root b is no instance of P as well. Assuming the contrary implies the existence of a homomorphism h from (the ABox translation of) P to $\exists Y.\mathcal{B}$ at b. In exactly the same manner as above for the case of an existential restriction, we can prove that $h(p) = a$ must hold true. This in turn would require that a is an instance of A—a contradiction, since $\mathsf{target}(p) \setminus A$ is not subsumed by A. Thus, b is not an instance of P if $p \neq P$. □

The two conditions stated in Lemma 10 are not only necessary, but also sufficient for safety for a singleton policy. Before we can prove this, we need the following auxiliary lemma.

Lemma 11. *Consider two ABoxes $\exists X.\mathcal{A}$ and $\exists Y.\mathcal{B}$. If $\mathcal{A} \cup \mathcal{B} \models C(u)$ for some object $u \in \Sigma_\mathsf{I} \cup Y$, but $\mathcal{B}$ does not entail $C(u)$, then there exists some path $p \in \mathsf{Paths}(C)$ and some individual name $a \in \Sigma_\mathsf{I}$ such that $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B} \models \mathsf{target}(p)(a)$ and $\exists Y.\mathcal{B} \not\models \mathsf{target}(p)(a)$.*

Proof. We prove the claim by induction on the role depth of C.

If u is an individual name, then the proof is already finished: set $p \coloneqq C$ and $a \coloneqq u$. Otherwise, u is a variable of the ABox $\exists Y.\mathcal{B}$. Since $\mathcal{B}$ does not entail $C(u)$, there must be some top-level conjunct $D \in \mathsf{Conj}(C)$ such that the concept assertion $D(u)$ is not entailed by $\mathcal{B}$. Furthermore, the precondition $\mathcal{A} \cup \mathcal{B} \models C(u)$ implies that $D(u)$ is entailed by $\mathcal{A} \cup \mathcal{B}$.

If $D = A$ is a concept name, then Lemma 2 implies that the concept assertion $A(u)$ is no element of $\mathcal{B}$ but is an element of $\mathcal{A} \cup \mathcal{B}$. It follows that $\mathcal{A}$ contains

$A(u)$—a contradiction, since $u \in Y$ and $(\Sigma_\mathsf{I} \cup X) \cap Y = \emptyset$. For the induction base, we infer that u cannot be a variable; otherwise the top-level conjunct D must be an existential restriction $\exists r. E$ and we proceed as follows.

From $\mathcal{A} \cup \mathcal{B} \models \exists r. E(u)$ we infer by an application of Lemma 2 that $\mathcal{A} \cup \mathcal{B}$ contains some role assertion $r(u, v)$ such that $\mathcal{A} \cup \mathcal{B}$ entails $E(v)$. Note that v is an object in $\Sigma_\mathsf{I} \cup Y$. An application of the induction hypothesis yields some path $q \in \mathsf{Paths}(E)$ and some individual name $a \in \Sigma_\mathsf{I}$ such that $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B} \models \mathsf{target}(q)(a)$ and $\exists Y. \mathcal{B} \not\models \mathsf{target}(q)(a)$. Since $p \coloneqq C \xrightarrow{r} q$ is a path in C such that $\mathsf{target}(p) = \mathsf{target}(q)$, we are done. □

Theorem 12. *$\exists X. \mathcal{A}$ is safe for $\{P\}$ iff the following two conditions are satisfied:*

1. *For each individual name a and for each concept name $A \in \mathsf{Atoms}(P)$, the concept assertion $A(a)$ is not in $\mathcal{A}$.*
2. *For each individual name a, for each role assertion $r(a, u)$ in $\mathcal{A}$, and for each existential restriction $\exists r. C$ in $\mathsf{Atoms}(P)$, there is no partial homomorphism from C to $\exists X. \mathcal{A}$ at u.*

Proof. The only-if direction has been shown in Lemma 10. We now turn our attention to proving the if direction.

Consider some ABox $\exists X. \mathcal{A}$ that is not safe for $\{P\}$. Non-safety implies that there exists an ABox $\exists Y. \mathcal{B}$ that is compliant with $\{P\}$, but for which the union with $\exists X. \mathcal{A}$ does not comply with $\{P\}$. We infer that there is some individual name $a \in \Sigma_\mathsf{I}$ such that $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B} \models P(a)$.

We now construct a sequence $((p_0, a_0), (p_1, a_1), \dots)$ with the following properties.

- $p_n \in \mathsf{Paths}(P)$, and $a_n \in \Sigma_\mathsf{I}$
- $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B} \models \mathsf{target}(p_n)(a_n)$
- $\exists Y. \mathcal{B} \not\models \mathsf{target}(p_n)(a_n)$
- $\mathsf{rd}(\mathsf{target}(p_n)) > \mathsf{rd}(\mathsf{target}(p_{n+1}))$

The sequence starts with $(p_0, a_0) \coloneqq (P, a)$ where P and a are from the last paragraph. The sequence is extended in the following way.

Let (p_n, a_n) be the last triple that has been defined. There must be some top-level conjunct $C \in \mathsf{Conj}(\mathsf{target}(p_n))$ such that $\exists Y. \mathcal{B} \not\models C(a_n)$, since $\exists Y. \mathcal{B} \not\models \mathsf{target}(p_n)(a_n)$ holds true. Note that C is an element of $\mathsf{Atoms}(P)$. Further note that $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B} \models \mathsf{target}(p_n)(a_n)$ implies $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B} \models C(a_n)$.

1. If $C = A$ is a concept name, then two applications of Lemma 2 yield that the concept assertion $A(a_n)$ cannot be in $\mathcal{B}$ but must be in the union $\mathcal{A} \cup \mathcal{B}$—we conclude that $\mathcal{A}$ contains $A(a_n)$ and the proof is finished.
2. In the remaining case C must be an existential restriction $\exists r. D$. Recall that the union $\exists X. \mathcal{A} \cup \exists Y. \mathcal{B}$ entails $\exists r. D(a_n)$. According to Lemma 2,

there exists a role assertion $r(a_n, u)$ in $\mathcal{A} \cup \mathcal{B}$ such that $D(u)$ is entailed by $\mathcal{A} \cup \mathcal{B}$. We proceed with a case distinction.

(a) If this role assertion $r(a_n, u)$ is in $\mathcal{A}$, then u is an object in $\Sigma_{\mathsf{I}} \cup X$. Since $\mathcal{A} \cup \mathcal{B} \models D(u)$, there is a homomorphism h from (the ABox translation of) D to the union $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B}$ at u. We now define the partial homomorphism $(j, \mathfrak{B})$ from D to $\exists X.\mathcal{A}$ at u where

$$\mathfrak{B} := \left\{ q \;\middle|\; \begin{array}{l} q \in \mathsf{Paths}(D) \text{ where } h(q) \in \Sigma_{\mathsf{I}} \\ \text{and } h(q') \in X \text{ for each prefix } q' \lneq q \end{array} \right\}$$
$$\cup \left\{ q \;\middle|\; \begin{array}{l} q \in \mathsf{Max}_{\leq}(\mathsf{Paths}(D)) \text{ where } h(q) \in X \\ \text{and } h(q') \in X \text{ for each prefix } q' \lneq q \end{array} \right\}$$

and where $j := h\restriction_{\downarrow\mathfrak{B}}$ is the restriction of the homomorphism h to the ideal $\downarrow\mathfrak{B}$. It is easy to see that $\mathfrak{B}$ is a maximal antichain in $(\mathsf{Paths}(D), \leq)$. The proof is finished.

(b) Otherwise, the role assertion $r(a_n, u)$ must be in $\mathcal{B}$, and u is an object in $\Sigma_{\mathsf{I}} \cup Y$. First note that $\mathcal{B}$ cannot entail $D(u)$, since $\exists Y.\mathcal{B}$ does not entail $\exists r.D(a_n)$. We further know that $\mathcal{A} \cup \mathcal{B}$ entails $D(u)$. Lemma 11 shows that there is some path $q \in \mathsf{Paths}(D)$ and some individual name $a' \in \Sigma_{\mathsf{I}}$ such that the union $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B}$ entails $\mathsf{target}(q)(a')$ but the ABox $\exists Y.\mathcal{B}$ alone does not entail $\mathsf{target}(q)(a')$. Note that $p_n \in \mathsf{Paths}(P)$, $\exists r.D \in \mathsf{Conj}(\mathsf{target}(p_n))$, and $q \in \mathsf{Paths}(D)$ implies that $p_n \xrightarrow{r} q$ is a path in P as well and it holds true that $\mathsf{target}(p_n \xrightarrow{r} q) = \mathsf{target}(q)$. Now extend the sequence with $(p_{n+1}, a_{n+1}) := (p_n \xrightarrow{r} q, a')$.

By construction of the sequence, the condition $\mathsf{rd}(\mathsf{target}(p_n)) > \mathsf{rd}(\mathsf{target}(p_{n+1}))$ is obviously satisfied for each index n except the last one. We conclude that the sequence must be finite, i.e., it cannot be extended after a finite number of elements. Of course, this can only be true if Case 1 or Case 2a above is satisfied after a finite number of iterations. □

Before using this characterization of safety to show that safety for singleton policies can be decided in polynomial time, let us apply it to the quantified ABoxes considered in the introduction. The quantified ABox in (2) clearly violates the first condition of the theorem since it contains the assertion *Comedian*(*JERRY*). The quantified ABox in (1) violates the second condition of the theorem since there is a partial homomorphism from $\mathit{Comedian} \sqcap \exists\, \mathit{spouse}.\, \mathit{Comedian}$ to the ABox at x. This can, for example, be seen by using the condition for the existence of a partial homomorphism given in Lemma 13 below. The existence of such a partial homomorphism crucially depends on the presence of the assertion *Comedian*(x). Since this assertion is missing in the quantified ABox in (3), this ABox satisfies both conditions of the theorem, and thus is safe.

Computational Complexity of Deciding Safety

First, we present a recursive characterization of existence of a partial homomorphism, and then show that this yields a polynomial time decision procedure for the existence problem.

Lemma 13. *There is a partial homomorphism from C to $\exists X.\, \mathcal{A}$ at u iff one of the following two statements is satisfied:*

1. *u is an individual name.*
2. *u is a variable and the following two statements are true:*
   - *(a) For each concept name $A \in \mathsf{Conj}(C)$, the matrix $\mathcal{A}$ contains the concept assertion $A(u)$.*
   - *(b) For each existential restriction $\exists r.\, D \in \mathsf{Conj}(C)$, the matrix $\mathcal{A}$ contains a role assertion $r(u, v)$ such that there is a partial homomorphism from D to $\exists X.\, \mathcal{A}$ at v.*

Proof. We begin with a proof of the *only if* direction. Let $(j, \mathfrak{B})$ be a partial homomorphism from C to $\exists X.\, \mathcal{A}$ at u. If u is an individual name, we are done.

Otherwise, consider the case where u is a variable. Then Definition 8 shows that $\mathcal{A}$ contains $A(u)$ for each concept name $A \in \mathsf{Conj}(C)$, since $j(C) = u$. Furthermore, Definition 8 implies that, for each existential restriction $\exists r.\, D \in \mathsf{Conj}(C)$, the matrix $\mathcal{A}$ contains the role assertion $r(u, j(C \xrightarrow{r} D))$. It further follows that the restriction $(j\restriction_D, \mathfrak{B}\restriction_D)$ where

$$\mathfrak{B}\restriction_D := \{\, p \mid p \in \mathsf{Paths}(D) \text{ and } C \xrightarrow{r} p \in \mathfrak{B} \,\}$$

and $j\restriction_D(q) := j(C \xrightarrow{r} q)$ for each $q \in \downarrow\mathfrak{B}\restriction_D$ is a partial homomorphism from D to $\exists X.\, \mathcal{A}$ at $j(C \xrightarrow{r} D)$.

It remains to show the *if* direction. If u is an individual name, then the pair $(j, \mathfrak{B})$ is a partial homomorphism where the maximal antichain is defined as $\mathfrak{B} := \{C\}$ and where the mapping $j\colon \downarrow\mathfrak{B} \to \Sigma_\mathsf{I} \cup X$ is defined by $j(C) := u$.

Otherwise, we have the case where u is a variable and the above Statements 2a and 2b are true. It then follows that the matrix $\mathcal{A}$ contains a subset

$$\{\, A(u) \mid A \in \mathsf{Conj}(C) \,\} \cup \{\, r(u, v_{\exists r.D}) \mid \exists r.\, D \in \mathsf{Conj}(C) \,\}$$

such that, for each $\exists r.\, D \in \mathsf{Conj}(C)$, there is a partial homomorphism $(j_{\exists r.D}, \mathfrak{B}_{\exists r.D})$ from D to $\exists X.\, \mathcal{A}$ at $v_{\exists r.D}$.

We construct a partial homomorphism $(j, \mathfrak{B})$ from C to $\exists X.\, \mathcal{A}$ at u as a union of the partial homomorphisms $(j_{\exists r.D}, \mathfrak{B}_{\exists r.D})$: the maximal antichain is defined as

$$\mathfrak{B} := \{\, C \xrightarrow{r} p \mid \exists r.\, D \in \mathsf{Conj}(C) \text{ and } p \in \mathfrak{B}_{\exists r.D} \,\}$$

and the mapping $j\colon {\downarrow}\mathfrak{B} \to \Sigma_\mathsf{I} \cup X$ is defined by $j(C) \coloneqq u$ and further

$$j(C \xrightarrow{r} p) \coloneqq j_{\exists r.D}(p)$$

for each $\exists r.D \in \mathsf{Conj}(C)$ and for each $p \in {\downarrow}\mathfrak{B}_{\exists r.D}$. The induced ideal of the border $\mathfrak{B}$ equals

$${\downarrow}\mathfrak{B} = \{C\} \cup \{\, C \xrightarrow{r} p \mid \exists r.D \in \mathsf{Conj}(C) \text{ and } p \in {\downarrow}\mathfrak{B}_{\exists r.D} \,\}.$$

We will now verify that the four conditions in Definition 8 are satisfied.

1. $j(C) = u$ is true by definition, see above.
2. Let $p \in {\downarrow}\mathfrak{B} \setminus \mathfrak{B}$, i.e., p is a path strictly below the border $\mathfrak{B}$. If $p = C$, then $j(p) = j(C) = u \in X$. Otherwise, $p = C \xrightarrow{r} q$ for some $q \in \mathsf{Paths}(D)$ and $\exists r.D \in \mathsf{Conj}(C)$ such that $q \in {\downarrow}\mathfrak{B}_{\exists r.D} \setminus \mathfrak{B}_{\exists r.D}$. We conclude that $j(p) = j(C \xrightarrow{r} q) = j_{\exists r.D}(q) \in X$.
3. Consider some $p \in \mathfrak{B} \setminus \mathsf{Max}_{\leq}(\mathsf{Paths}(C))$, i.e., p is in the border $\mathfrak{B}$ but is not a leaf. Then p is of the form $C \xrightarrow{r} q$ for some $q \in \mathfrak{B}_{\exists r.D}$ and $\exists r.D \in \mathsf{Conj}(C)$. Since p is no leaf (of C), it follows that q is not a leaf (of D) as well. We conclude that $j(p) = j(C \xrightarrow{r} q) = j_{\exists r.D}(q) \in \Sigma_\mathsf{I}$.
4. Assume that $p \in {\downarrow}\mathfrak{B}$, i.e., p is in or below the border $\mathfrak{B}$, and let $j(p) \in X$.
   (a) Consider a concept name $A \in \mathsf{Conj}(\mathsf{target}(p))$. We show that $A(j(p)) \in \mathcal{A}$. In the case $p = C$ we have $j(p) = u$ and $A \in \mathsf{Conj}(C)$, and we have already concluded above that $A(u)$ is contained in $\mathcal{A}$.

   Otherwise, p must be of the form $C \xrightarrow{r} q$ where $q \in {\downarrow}\mathfrak{B}_{\exists r.D}$ and $\exists r.D \in \mathsf{Conj}(C)$. It then follows that $j(p) = j(C \xrightarrow{r} q) = j_{\exists r.D}(q)$, which yields $j_{\exists r.D}(q) \in X$ and thus $A(j_{\exists r.D}(q)) \in \mathcal{A}$, i.e., $A(j(p)) \in \mathcal{A}$.

   (b) Let $\exists s.E \in \mathsf{Conj}(\mathsf{target}(p))$. We will show that $s(j(p), j(p \xrightarrow{s} E)) \in \mathcal{A}$. If $p = C$, then $j(p) = u$ and $\exists s.E \in \mathsf{Conj}(C)$. We already know that $\mathcal{A}$ contains a role assertion $s(u, v_{\exists s.E})$ where $v_{\exists s.E} = j_{\exists s.E}(E) = j(C \xrightarrow{s} E)$.

   In the remaining case we have $p = C \xrightarrow{r} q$ where $q \in {\downarrow}\mathfrak{B}_{\exists r.D}$ and $\exists r.D \in \mathsf{Conj}(C)$. Note that $\mathsf{target}(q) = \mathsf{target}(p)$. Then $j(p) = j(C \xrightarrow{r} q) = j_{\exists r.D}(q)$. It follows that $j_{\exists r.D}(q) \in X$, and so we have $s(j_{\exists r.D}(q), j_{\exists r.D}(q \xrightarrow{s} E)) \in \mathcal{A}$, which implies $s(j(p), j(p \xrightarrow{s} E)) \in \mathcal{A}$. □

Proposition 14. *It can be decided in polynomial time whether there exists a partial homomorphism from C to $\exists X.\mathcal{A}$ at u.*

Proof. We show the claim by induction on the role depth of C. It takes linear time to check whether u is an individual. If so, we can immediately return an affirmative answer. Otherwise, if u is a variable, we need to check whether the matrix $\mathcal{A}$ contains the concept assertion $A(u)$ for each concept name A in the top-level conjunction of C, which can clearly be done in polynomial time. For the base case, where C only contains concept names, we are already done, and just answer affirmatively if all the aforementioned tests succeed, and answer negatively otherwise.

For the step case, Lemma 13 tells us that we further need to check if, for each existential restriction $\exists r.\, D$ in $\mathsf{Conj}(C)$, there is a role assertion $r(u, v)$ in the matrix $\mathcal{A}$ such that there is a partial homomorphism from D to $\exists X.\, \mathcal{A}$ at v. Of course, there are at most polynomially many r-successors v of u and, for each of them, the induction hypothesis implies that we can decide existence of a partial homomorphism from D to $\exists X.\, \mathcal{A}$ at v in polynomial time. Thus, all required tests can be conducted in polynomial time. □

The following result is now an immediate consequence of this proposition and Theorem 12.

Corollary 15. *It can be decided in polynomial time if a quantified ABox is safe for a singleton policy.*

How to Deal with Non-Singleton Policies

To start with, let us ask whether general policies are indeed more expressive than singleton policies. The following example answers this question in the affirmative, by showing that not every policy is safety-equivalent to a singleton policy. Here *safety-equivalent* means that the same ABoxes are safe for the two policies.

Example 16. Consider the policy $\mathcal{P} \coloneqq \{A,\, \exists r.\, (A \sqcap B)\}$, and assume that there is a singleton policy $\{P\}$ such that $\mathcal{P}$ is safety-equivalent to $\{P\}$.

It is easy to see that the quantified ABox $\exists \emptyset.\, \{r(a, b),\, B(b)\}$ is $\mathcal{P}$-safe. Thus, it must be $\{P\}$-safe as well. According to Theorem 12, this implies that $\mathsf{Atoms}(P)$ cannot contain an existential restriction of the form $\exists r.\, C$.

We claim that this implies that the quantified ABox $\exists \{x\}.\, \{r(a, x),\, A(x),\, B(x)\}$ is safe for $\{P\}$. This yields a contradiction to our assumption that $\mathcal{P}$ is safety-equivalent to $\{P\}$ since this quantified ABox is not even compliant with $\mathcal{P}$.

To prove the claim, we use again Theorem 12. Since the quantified ABox does not contain a concept assertion for an individual name, the first condition of the theorem is satified. The second condition is satisfied as well since $\mathsf{Atoms}(P)$ does not contain an existential restriction for the role r.

Our characterization of safety for the case of singleton policies cannot be extended in a straightforward way to general policies $\mathcal{P}$. The main problem appears to be that the constructions of counterexample ABoxes employed above need not yield compliant ABoxes. The next example shows that non-compliance with

$\mathsf{Atoms}(\mathcal{P}) := \bigcup\{\, \mathsf{Atoms}(P) \mid P \in \mathcal{P} \,\}$ does not necessarily lead to a violation of safety.

Example 17. Consider the policy $\mathcal{P} := \{B \sqcap \exists r.\,(A_1 \sqcap A_2),\ A_1\}$. The ABox $\exists\emptyset.\,\{A_2(a)\}$ is easily seen to be safe for $\mathcal{P}$, although it is not compliant with $\mathsf{Atoms}(\mathcal{P})$ since $A_2 \in \mathsf{Atoms}(\mathcal{P})$. If we had the singleton policy $\{B \sqcap \exists r.\,(A_1 \sqcap A_2)\}$, then our construction would yield the ABox $\exists\emptyset.\,\{B(b),\ r(b,a),\ A_1(a)\}$ as counterexample to safety. However, since $A_1 \in \mathcal{P}$, this ABox is not compliant with $\mathcal{P}$.

A possible approach for preventing this problem is to restrict attention to the subset $\mathsf{SafetyAtoms}(\mathcal{P})$ of $\mathsf{Atoms}(\mathcal{P})$ consisting of all atoms C that are a top-level conjunct of $\mathsf{target}(p)$ for some path p in a policy concept $P \in \mathcal{P}$, but for which $\mathsf{target}(p) \not\sqsubseteq_\emptyset Q$ for each $Q \in \mathcal{P} \setminus \{P\}$. If we replace $\mathsf{Atoms}(\mathcal{P})$ with $\mathsf{SafetyAtoms}(\mathcal{P})$, then Lemma 6 also holds for non-singleton policies.

Even with this modification, Lemma 10 needs no longer hold. To see this, consider Figure 4. The small gray triangles in this figure remain in the constructed ABox, with an individual name at the root. Thus, the corresponding subconcepts $\mathsf{target}(p \xrightarrow{r} q_j)$ should not be subsumed by any policy concept since otherwise the constructed ABox cannot be compliant. The following example shows that, even if we impose this restriction in the definition of a partial homomorphism, Lemma 10 still does not hold.

Example 18. Consider the policy $\mathcal{P} := \{\exists r.\,(\exists r.\,A \sqcap \exists r.\,B),\ A \sqcap B\}$ and the ABox $\exists X.\,\mathcal{A} := \exists\{x\}.\,\{r(a,x),\ r(x,b)\}$, which can easily be seen to be safe. There is a partial homomorphism from $\exists r.\,A \sqcap \exists r.\,B$ to the ABox at x, namely $(j, \{\exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} A,\ \exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} B\})$ where $j(\exists r.\,A \sqcap \exists r.\,B) := x$ and $j(\exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} A) := b$ and $j(\exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} B) := b$. Neither $\mathsf{target}(\exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} A)$ nor $\mathsf{target}(\exists r.\,A \sqcap \exists r.\,B \xrightarrow{r} B)$ is subsumed by a policy concept, but their conjunction is subsumed by $A \sqcap B$, i.e., the constructed ABox cannot be compliant. In particular, $\exists Y.\,\mathcal{B}$ looks as follows: $\exists\emptyset.\,\{A(a),\ A(b),\ B(b)\}$. It entails $(A \sqcap B)(b)$.

At the moment, we do not have a characterization of safety for the case of non-singleton policies that is in the spirit of Theorem 12. Nevertheless, using ideas from [12, 13] it is easy to see that safety for general policies is in NP.

Proposition 19. *Safety for general policies can be decided in nondeterministic polynomial time.*

Proof Sketch. The main idea underlying the proof is that, whenever $\exists X.\,\mathcal{A}$ is not safe for $\mathcal{P}$, then there exists a *small* ABox $\exists Y.\,\mathcal{B}$ that is compliant with $\mathcal{P}$ and such that $\exists X.\,\mathcal{A} \cup \exists Y.\,\mathcal{B}$ is not compliant with $\mathcal{P}$, where small means that the number of object names occurring in $\mathcal{B}$ is polynomially bounded by the maximal size of the concepts in $\mathcal{P}$. Such an ABox can then be guessed in nondeterministic polynomial time.

The reason for the existence of such a small counterexample to safety is the following. If $\exists X.\mathcal{A}$ is not safe for $\mathcal{P}$, then there exists a compliant quantified ABox $\exists Z.\mathcal{C}$ such that $\exists X.\mathcal{A} \cup \exists Z.\mathcal{C} \models P(a)$ for an individual a and a policy concept $P \in \mathcal{P}$. Thus, there is a homomorphism from the ABox translation of $P(a)$ to $\exists X.\mathcal{A} \cup \exists Z.\mathcal{C}$. Let $\exists Y.\mathcal{B}$ be the quantified ABox obtained from $\exists Z.\mathcal{C}$ by removing all objects that are not in the image of this homomorphism. This provides us with the small ABox we are looking for. □

4 The Optimal Safe Anonymization

If a given quantified ABox turns out not to be safe, we want to modify it in a minimal way to make it safe before publishing it. Given a quantified ABox $\exists X.\mathcal{A}$ and a policy $\mathcal{P}$, we say that $\exists Y.\mathcal{B}$ is a *$\mathcal{P}$-compliant anonymization* (*$\mathcal{P}$-safe anonymization*) of $\exists X.\mathcal{A}$ if $\exists X.\mathcal{A} \models \exists Y.\mathcal{B}$ and $\exists Y.\mathcal{B}$ is compliant with $\mathcal{P}$ (safe for $\mathcal{P}$). Such an anonymization $\exists Y.\mathcal{B}$ is *optimal* if there is no $\mathcal{P}$-compliant anonymization ($\mathcal{P}$-safe anonymization) of $\exists X.\mathcal{A}$ that lies strictly between $\exists X.\mathcal{A}$ and $\exists Y.\mathcal{B}$ w.r.t. the entailment relation. Thus, optimality means that we minimize the amount of entailments lost by the anonymization.

The problem of computing optimal $\mathcal{P}$-compliant anonymizations of quantified ABoxes for $\mathcal{EL}$ policies was investigated in detail in [3], where it is shown that a quantified ABox may in the worst case have exponentially many $\mathcal{P}$-compliant anonymizations of exponential size. We show below that, for safety w.r.t. singleton policies, there always exists an (up to equivalence) *unique* optimal anonymization, which may, however, still be of exponential size.

Our construction of this unique safe anonymization is inspired by the approach employed in [3] for the case of compliance. The main idea underlying that approach is that one needs to generate copies of objects, rather than just remove assertions. For example, consider the quantified ABox $\exists \{x\}. \{r(a,x), A_1(x), A_2(x), A_3(x)\}$ and the policy concept $P \coloneqq \exists r.(A_1 \sqcap A_2 \sqcap A_3)$. Compliance can, e.g., be achieved by removing $A_1(x)$, but the resulting ABox is not optimal. In fact, one can obtain an optimal compliant anonymization by introducing three copies y_1, y_2, y_3 of x, making all of them variables and r-successors of a, and adding for all $i, 1 \leq i \leq 3$, the assertions $A_k(y_i)$ and $A_\ell(y_i)$ where $\{k, \ell\} = \{1,2,3\} \setminus \{i\}$. In the general construction, the copies of an object name u occurring in $\exists X.\mathcal{A}$ are basically of the form $y_{u,\mathcal{K}}$ where $\mathcal{K} \subseteq \mathsf{Atoms}(\mathcal{P})$. The variables y_i in our example would actually be denoted by $y_{x,\{A_i\}}$ in this construction. The quantified ABox $\exists Y.\mathcal{B}$ containing these copies is then defined in a way which ensures that

- $\exists Y.\mathcal{B}$ does not entail $C(y_{u,\mathcal{K}})$ if $C \in \mathcal{K}$.

A so-called compliance seed function determines which copy of an individual a is employed to represent this individual. It is defined in a way that ensures compliance (see [3] for details). In our example, the seed function uses $y_{a,\{\exists r.(A_1 \sqcap A_2 \sqcap A_3)\}}$ to represent a.

Inspired by this idea, we also employ such copies $y_{u,\mathcal{K}}$ in our construction of the optimal safe anonymization $\exists Y.\mathcal{B}$. However, we view all such copies as variables, and explicitly keep the individual names from $\exists X.\mathcal{A}$ to denote individuals. The intuition underlying the sets $\mathcal{K}$ also differs from the one in the case of compliance. In fact, the ABox $\exists Y.\mathcal{B}$ is constructed such that the following holds:

- if $y_{u,\mathcal{K}}$ is a variable in $\exists Y.\mathcal{B}$ and $C \in \mathcal{K}$, then there is no partial homomorphism from C to $\exists Y.\mathcal{B}$ at $y_{u,\mathcal{K}}$.

Given the close connection between the entailment of concept assertions and the existence of homomorphisms, this condition actually modifies the one used in the case of compliance by replacing "homomorphism" with "partial homomorphism."

Before defining the optimal safe anonymization of $\exists X.\mathcal{A}$ formally, we introduce an optimization (also employed in [3]) that allows us to reduce the number of copies $y_{u,\mathcal{K}}$ that must be introduced. This optimization is based on the following lemma.

Lemma 20. *Let C, D be $\mathcal{EL}$ concept descriptions and $\exists X.\mathcal{A}$ a quantified ABox. If there is a partial homomorphism from C to $\exists X.\mathcal{A}$ at u and $C \sqsubseteq_\emptyset D$, then there also is a partial homomorphism from D to $\exists X.\mathcal{A}$ at u.*

Proof. Assume that there is a partial homomorphism from C to $\exists X.\mathcal{A}$ at u and further that C is subsumed by D. According to Lemma 13, the first assumption implies that we need to distinguish two cases.

1. In the first case, u is an individual name. Lemma 13 immediately yields that there exists a partial homomorphism from D to $\exists X.\mathcal{A}$ at u as well.
2. In the second case, u is a variable and further the two Statements 2a and 2b in Lemma 13 are satisfied for C. To justify the existence of a partial homomorphism from D to $\exists X.\mathcal{A}$ at u we shall show that the two Statements 2a and 2b are also true for D.
   (a) Let $A \in \mathsf{Conj}(D)$. Due to $C \sqsubseteq_\emptyset D$, we also have that $A \in \mathsf{Conj}(C)$, cf. Lemma 1. Since Statement 2a is true for C, it follows that $\mathcal{A}$ contains the concept assertion $A(u)$. Thus, Statement 2a is satisfied for D as well.
   (b) Let $\exists r.F \in \mathsf{Conj}(D)$. Since $C \sqsubseteq_\emptyset D$, Lemma 1 yields some $\exists r.E \in \mathsf{Conj}(C)$ such that $E \sqsubseteq_\emptyset F$. As Statement 2b holds true for C, we infer that $\mathcal{A}$ contains some role assertion $r(u, v)$ such that there is a partial homomorphism from E to $\exists X.\mathcal{A}$ at v. Induction yields that

> there is a partial homomorphism from F to $\exists X.\mathcal{A}$ at v, which shows that Statement 2b holds for D as well. □

Consequently, if $C \sqsubseteq_\emptyset D$ and $D \in \mathcal{K}$ prevents the existence of a partial homomorphism from D to $\exists Y.\mathcal{B}$ at $y_{u,\mathcal{K}}$, then this also prevents the the existence of a partial homomorphism from C to $\exists Y.\mathcal{B}$ at $y_{u,\mathcal{K}}$. Thus, it is sufficient to have only the subsumer D in $\mathcal{K}$. This insight allows us to restrict the sets $\mathcal{K}$ to ones not containing any $\sqsubseteq_\emptyset$-comparable elements.

Definition 21. The *canonical safe anonymization* $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ of $\exists X.\mathcal{A}$ w.r.t. some singleton policy $\{P\}$ is the ABox $\exists Y.\mathcal{B}$ consisting of the following components. As set of variables, we use

$$Y := \left\{ y_{u,\mathcal{K}} \;\middle|\; \begin{array}{l} u \in \Sigma_\mathsf{I} \cup X,\ \mathcal{K} \subseteq \mathsf{Atoms}(P), \text{ and} \\ \mathcal{K} \text{ does not contain } \sqsubseteq_\emptyset\text{-comparable atoms} \end{array} \right\}.$$

The matrix $\mathcal{B}$ is then constructed as follows:

1. Add the concept assertion $A(a)$ to $\mathcal{B}$ if $A(a)$ is in $\mathcal{A}$ and $A \notin \mathsf{Atoms}(P)$.
2. Add the concept assertion $A(y_{u,\mathcal{K}})$ to $\mathcal{B}$ if $A(u)$ is in $\mathcal{A}$ and $A \notin \mathcal{K}$.
3. Add the role assertion $r(a,b)$ to $\mathcal{B}$ if $r(a,b)$ is in $\mathcal{A}$ and there is no existential restriction $\exists r.C \in \mathsf{Atoms}(P)$.
4. Add the role assertion $r(a, y_{v,\mathcal{L}})$ to $\mathcal{B}$ if $r(a,v)$ is in $\mathcal{A}$ and, for each existential restriction $\exists r.C \in \mathsf{Atoms}(P)$, the set $\mathcal{L}$ contains some atom subsuming C, i.e., there is some $D \in \mathcal{L}$ such that $C \sqsubseteq_\emptyset D$.
5. Add the role assertion $r(y_{u,\mathcal{K}}, y_{v,\mathcal{L}})$ to $\mathcal{B}$ if $r(u,v)$ is in $\mathcal{A}$ and, for each existential restriction $\exists r.C \in \mathcal{K}$, the set $\mathcal{L}$ contains some atom subsuming C, i.e., there is some $D \in \mathcal{L}$ such that $C \sqsubseteq_\emptyset D$.
6. Add the role assertion $r(y_{u,\mathcal{K}}, b)$ to $\mathcal{B}$ if $r(u,b)$ is in $\mathcal{A}$ and there is no existential restriction $\exists r.C \in \mathcal{K}$.

Note that no role assertion is added to the matrix $\mathcal{B}$ for the case $C = \top$ in the above Statements 4 and 5, as there are no atoms subsuming $\top$.

In the remainder of this section, we show that $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ is indeed the optimal $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$, and that it can be computed in exponential time.

First, note that (2), (5), and (6) of the construction together with Lemma 13 ensure that the intuition underlying the variables $y_{u,\mathcal{K}}$ mentioned above is really satisfied by $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.

Lemma 22. *If C is a concept description and $y_{u,\mathcal{K}}$ is a variable such that $\mathcal{K}$ contains some atom D with $C \sqsubseteq_\emptyset D$, then there is no partial homomorphism from C to $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ at $y_{u,\mathcal{K}}$.*

Proof. We show the claim by induction on the role depth of C. If the atom D is a concept name A, then $A \in \mathsf{Conj}(C)$ and $A(y_{v,\mathcal{L}})$ is not in $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$. Lemma 13 yields the claim.

Otherwise, D is an existential restriction $\exists s.\,E$, i.e., there is some $\exists s.\,E' \in \mathsf{Conj}(C)$ such that $E' \sqsubseteq_\emptyset E$. For each individual name b that is an s-successor of $y_{v,\mathcal{L}}$, Statement 6 tells us that there does not exist an existential restriction $\exists s.\,E''$ in $\mathcal{L}$. We conclude that $y_{v,\mathcal{L}}$ does not have individual names as s-successors.

Consider a variable $y_{w,\mathcal{M}}$ that is an s-successor of $y_{v,\mathcal{L}}$. Statement 5 implies that $\mathcal{M}$ contains some atom F such that $E \sqsubseteq_\emptyset F$. Now the induction hypothesis immediately yields that there cannot exist a partial homomorphism from E to $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ at $y_{w,\mathcal{M}}$.

As there is no partial homomorphism from E to $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ at any s-successor of $y_{v,\mathcal{L}}$, there cannot exist a compliant partial homomorphism from E' at any s-successor of $y_{v,\mathcal{L}}$ either, cf. Lemma 20. Finally, an application of Lemma 13 shows the claim. □

This lemma, together with the characterization of safety given in Theorem 12 and (1), (3), and (4) of the construction, then yields that $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ is indeed safe for $\{P\}$.

Proposition 23. *The quantified ABox* $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ *is entailed by* $\exists X.\,\mathcal{A}$ *and safe for* $\{P\}$*.*

Proof. As an easy consequence of Definition 21 we obtain that the mapping h where $h(a) \coloneqq a$ for each individual name a and where $h(y_{u,\mathcal{K}}) \coloneqq u$ for each variable $y_{u,\mathcal{K}}$ is a homomorphism from $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ to $\exists X.\,\mathcal{A}$. This shows that $\exists X.\,\mathcal{A}$ entails $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$.

We make use of Theorem 12 for justifying safety. Consider an individual name a and a concept name $A \in \mathsf{Atoms}(P)$. By the very definition of $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$, its matrix $\mathcal{B}$ does not contain the concept assertion $A(a)$.

It remains to prove that, for each individual name a, for each role assertion $r(a, u)$ in the matrix of $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$, and for each existential restriction $\exists r.\,C$ in $\mathsf{Atoms}(P)$, there does not exist a partial homomorphism from C to $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ at u. Note that Lemma 20 tells us that it suffices to consider existential restrictions $\exists r.\,C$ in $\mathsf{Max}(\mathsf{Atoms}(P))$.

If u is an individual name, then by (3) of Definition 21 there is no existential restriction $\exists r.\,C$ in $\mathsf{Max}(\mathsf{Atoms}(P))$. Thus, there is nothing to show. Now assume that u is a variable $y_{v,\mathcal{L}}$. Since $r(a, y_{v,\mathcal{L}})$ is a role assertion in $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$, (4) of Definition 21 implies that $\mathcal{A}$ contains $r(a, v)$ and that $C \sqsubseteq_\emptyset D$ for some atom $D \in \mathcal{L}$. Lemma 22 yields that there is no partial homomorphism from C to $\mathsf{sa}(\exists X.\,\mathcal{A}, \{P\})$ at $y_{v,\mathcal{L}}$. □

The following proposition implies optimality of $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.

Proposition 24. *Each $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$ is entailed by* $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.

Proof. Let $\exists Z.\mathcal{C}$ be a $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$. Then there is a homomorphism h from $\exists Z.\mathcal{C}$ to $\exists X.\mathcal{A}$. Define the mapping k by setting $k(a) \coloneqq a$ for each individual name a and $k(x) \coloneqq y_{h(x),f(x)}$ for each variable x where

$$\begin{aligned} f(x) &\coloneqq \{\, A \mid A \in \mathsf{Atoms}(P) \text{ and } A(x) \notin \mathcal{C} \,\} \\ &\quad \cup \mathsf{Max}\left\{ \exists r.C \;\middle|\; \begin{array}{l} \exists r.C \in \mathsf{Atoms}(P) \text{ and for each } r(x,u) \in \mathcal{C}, \\ \text{there is no partial homomorphism} \\ \text{from } C \text{ to } \exists Z.\mathcal{C} \text{ at } u \end{array} \right\}. \end{aligned}$$

We prove that k is a homomorphism from $\exists Z.\mathcal{C}$ to $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.

1. Let $A(a) \in \mathcal{C}$, which implies $A(a) \in \mathcal{A}$. Since $\exists Z.\mathcal{C}$ is safe for $\mathcal{P}$, Lemma 10 implies that A cannot be contained in $\mathsf{Atoms}(P)$, and so (1) of Definition 21 ensures that the concept assertion $A(a)$ is contained in $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.
2. Let $A(x) \in \mathcal{C}$, which implies $A(h(x)) \in \mathcal{A}$. It follows that $A \notin f(x)$ and so we conclude by (2) of Definition 21 that $A(k(x))$ is in $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.
3. Consider a role assertion $r(a,b) \in \mathcal{C}$, which then also belongs to $\mathcal{A}$. Since $\exists Z.\mathcal{C}$ is safe for $\mathcal{P}$, Lemma 10 implies that, for each $\exists r.C \in \mathsf{Max}(\mathsf{Atoms}(P))$, there is no partial homomorphism from C to $\exists Z.\mathcal{C}$ at b. Since b is an individual name, Lemma 13 implies that, for each $\exists r.C \in \mathsf{Max}(\mathsf{Atoms}(P))$, there always exists a partial homomorphism from C to $\exists Z.\mathcal{C}$ at b. We conclude that $\mathsf{Max}(\mathsf{Atoms}(P))$ cannot contain an existential restriction $\exists r.C$. Thus (3) of Definition 21 yields that $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ contains $r(a,b)$.
4. Let $r(a,y)$ be a role assertion in $\mathcal{C}$, and thus $r(a,h(y)) \in \mathcal{A}$, and let $\exists r.C \in \mathsf{Max}(\mathsf{Atoms}(P))$. According to Lemma 10, there does not exist a partial homomorphism from C to $\exists Z.\mathcal{C}$ at y. Since y is a variable, Lemma 13 implies that either there is a concept name $A \in \mathsf{Conj}(C)$ such that the concept assertion $A(y)$ is not in $\mathcal{C}$, or there is an existential restriction $\exists s.D \in \mathsf{Conj}(C)$ such that, for each $s(y,v) \in \mathcal{C}$, there is no partial homomorphism from D to $\exists Z.\mathcal{C}$ at v. In the first case, A is in $f(y)$. In the second case, Lemma 20 yields that $f(y)$ contains some atom subsuming $\exists s.D$. In both cases, we have that some atom in $f(y)$ subsumes C, and thus (4) ensures that the role assertion $r(a,k(y))$ is indeed in $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.
5. Let $r(x,y)$ in $\mathcal{C}$, which yields $r(h(x),h(y)) \in \mathcal{A}$. In addition, consider some existential restriction $\exists r.C$ in $f(x)$, i.e., there does not exist any partial homomorphism from C to $\exists Z.\mathcal{C}$ at y. Since y is a variable, Lemma 13 implies that either there is a concept name $A \in \mathsf{Conj}(C)$ such that the concept assertion $A(y)$ is not in $\mathcal{C}$, or there is an existential restriction $\exists s.D \in \mathsf{Conj}(C)$

such that, for each $s(y, v) \in \mathcal{C}$, there is no partial homomorphism from D to $\exists Z.\mathcal{C}$ at v. In the first case, A is in $f(y)$. In the second case, Lemma 20 yields that $f(y)$ contains some atom subsuming $\exists s.\, D$. In both cases, we have that some atom in $f(y)$ subsumes C, and thus (4) yields that the role assertion $r(k(x), k(y))$ is indeed in $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$.

6. Finally, let $r(x, b) \in \mathcal{C}$, and thus $r(h(x), b)$ is in $\mathcal{A}$. By definition of f we have that, for each $\exists r.\, C \in f(x)$, there does not exist any partial homomorphism from C to $\exists Z.\mathcal{C}$ at b. Since b is an individual name, Lemma 13 yields that, for each $\exists r.\, C \in f(x)$, there is always a partial homomorphism from C to $\exists Z.\mathcal{C}$ at b. We conclude that $f(x)$ cannot contain any existential restriction $\exists r.\, C$. Now (6) ensures that $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ contains $r(k(x), b)$. □

Putting the results of Propositions 23 and 24 together, we obtain:

Theorem 25. *The quantified ABox $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ is the (up to equivalence) unique optimal $\{P\}$-safe anonymization of $\exists X.\mathcal{A}$.*

The cardinality of the set $\mathsf{Atoms}(P)$ is linear in the size of P, and thus we need to create at most exponentially many copies of each object in $\exists X.\mathcal{A}$. In addition, the conditions for whether to include an assertion in the constructed ABox $\exists Y.\mathcal{B}$ can be tested in polynomial time. Thus, the above theorem yields the following complexity results.

Corollary 26. *The optimal $\{P\}$-safe anonymization $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ of $\exists X.\mathcal{A}$ can be computed in exponential time for combined complexity and in polynomial time for data complexity.*

A slight modification of Example 2 in [2] can be used to show that the exponential upper bound stated in the corollary is tight.

Example 27. Consider the ABox $\exists X.\mathcal{A}$ with variable x and matrix

$$\{r(a, x),\ A_1(x),\ B_1(x),\ \ldots,\ A_n(x),\ B_n(x)\},$$

and the policy concept $P := \exists r.\, (A_1 \sqcap B_1) \sqcap \ldots \sqcap \exists r.\, (A_n \sqcap B_n)$. The set $\mathsf{Atoms}(P)$ contains the concept names A_i and B_i as well as the existential restriction $\exists r.\, (A_i \sqcap B_i)$ for each index i. It is easy to see that the optimal safe anonymization $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ must contain exponentially many r-successors of the individual a, namely the variables $y_{x,\mathcal{K}}$ for each set $\mathcal{K}$ that contains either A_i or B_i for each index i.

Assume that there were a quantified ABox $\exists Z.\mathcal{C}$ that is equivalent to $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ but which does not have an exponential size. It would follow that, in $\exists Z.\mathcal{C}$, there must be an r-successor z of a such that at least two of the variables $y_{x,\mathcal{K}}$ in $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ are mapped to z (for a fixed homomorphism

h from $\mathsf{sa}(\exists X.\mathcal{A}, \{P\})$ to $\exists Z.\mathcal{C})$. We could conclude that z must be an instance of both A_i and B_i for some index i, which would imply that a is an instance of the atom $\exists r.(A_i \sqcap B_i)$—a contradiction to the second condition of Lemma 6.

In order to compute a safe anonymization of some quantified ABox $\exists X.\mathcal{A}$ w.r.t. a non-singleton policy $\{P_1, \ldots, P_k\}$, we could of course employ the above construction iteratively by considering the policy concepts P_i one after another, i.e., first compute the canonical safe anonymization of $\exists X.\mathcal{A}$ w.r.t. $\{P_1\}$, then compute the canonical safe anonymization of the result w.r.t. $\{P_2\}$, etc. While this iterated computation always yields a quantified ABox that is safe for the whole policy $\{P_1, \ldots, P_k\}$, it need not produce an optimal safe anonymization. A counterexample is as follows.

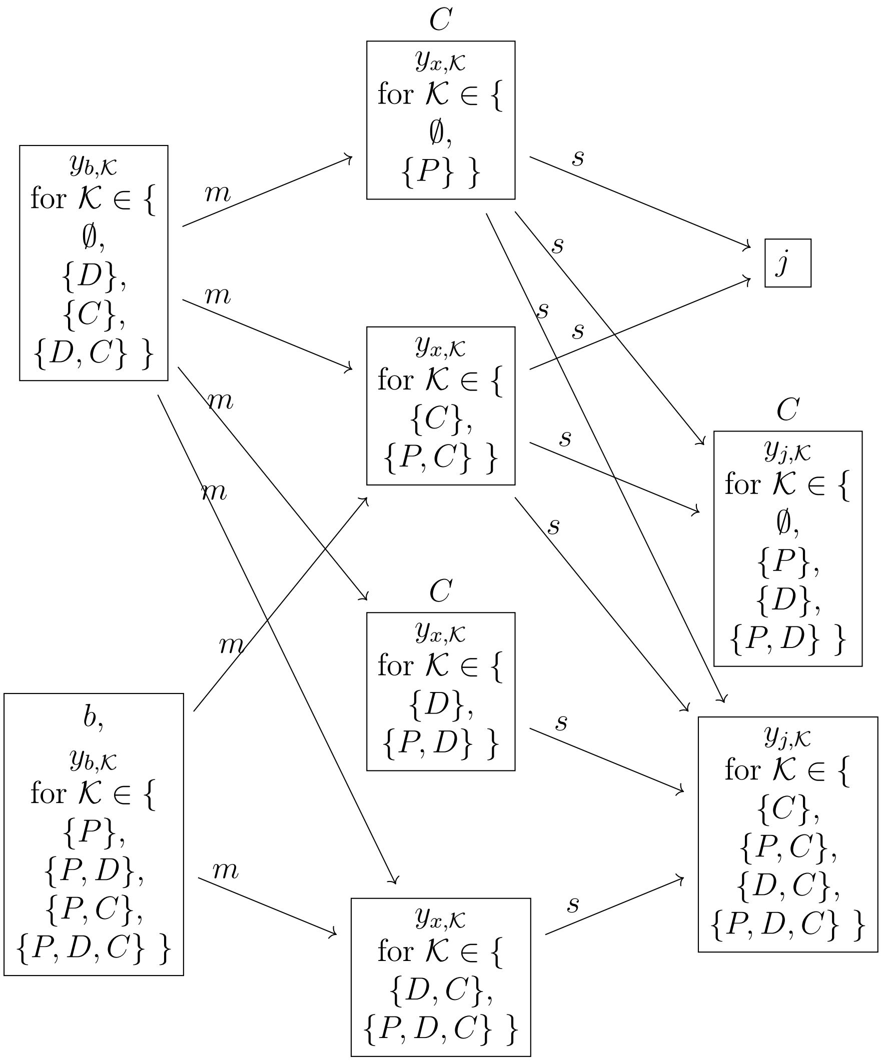
Example 28. Consider the quantified ABox $\exists\emptyset.\{r(a, b)\}$ and the policy $\{A, \exists r.A\}$. It is obvious that the ABox is safe, i.e., it is its own optimal safe anonymization. However, it is not safe for the singleton policy $\{\exists r.A\}$. Thus, the iterated approach would make the ABox safe for that singleton policy and so removes the assertion $r(a, b)$, destroying optimality.

Finally, let us come back to the Ben and Jerry example from the introduction. Figure 5 depicts the canonical safe anonymization of Ben's original ABox, where we use obvious abbreviations for concept, role, and individual names. This shows that the safe anonymization (3) we came up with in the introduction is not optimal. In fact, the canonical safe anonymization implies that Ben is an instance of the concept $\exists \mathit{mother}.\exists \mathit{spouse}.\mathit{Comedian}$, whereas (3) does not have this consequence.

5 Conclusion

We have shown that deciding safety of a quantified ABox w.r.t. a policy defined by a single $\mathcal{EL}$ concept can be decided in polynomial time, and that the unique optimal safe anonymization of a non-safe quantified ABox can be computed in exponential time. Both complexity results are w.r.t. combined complexity, where both the data and the policy are view to be part of the input. For data complexity (where the policy is assumed to be fixed), the complexity of the latter problem also drops to polynomial time. In the worst case, the exponential complexity for computing the optimal safe anonymization cannot be avoided, as demonstrated by Example 27.

Compared to the findings in [12, 13], our results show that the restriction from conjunctive queries to $\mathcal{EL}$ concepts as formalism for representing the policy pays off complexity-wise: in the setting considered in [12, 13], the complexity of deciding safety lies on the second level of the polynomial hierarchy. It would be


C
$y_{x,\mathcal{K}}$ for $\mathcal{K} \in \{ \emptyset, \{P\} \}$
$y_{b,\mathcal{K}}$ for $\mathcal{K} \in \{ \emptyset, \{D\}, \{C\}, \{D, C\} \}$
m
m
m
m
s
s
s
s
s
s
s
s
j
$y_{x,\mathcal{K}}$ for $\mathcal{K} \in \{ \{C\}, \{P, C\} \}$
C
$y_{j,\mathcal{K}}$ for $\mathcal{K} \in \{ \emptyset, \{P\}, \{D\}, \{P, D\} \}$
C
$y_{x,\mathcal{K}}$ for $\mathcal{K} \in \{ \{D\}, \{P, D\} \}$
m
b, $y_{b,\mathcal{K}}$ for $\mathcal{K} \in \{ \{P\}, \{P, D\}, \{P, C\}, \{P, D, C\} \}$
m
$y_{x,\mathcal{K}}$ for $\mathcal{K} \in \{ \{D, C\}, \{P, D, C\} \}$
$y_{j,\mathcal{K}}$ for $\mathcal{K} \in \{ \{C\}, \{P, C\}, \{D, C\}, \{P, D, C\} \}$

Original ABox: $\exists\{x\}.\{m(b, x),\ C(x),\ s(x, j),\ C(j)\}$
Policy: $\{\exists m.(C \sqcap \exists s.C)\}$
Abbreviations: $P \coloneqq \exists m.(C \sqcap \exists s.C)$ and $D \coloneqq \exists s.C$

Figure 5: The canonical safe anonymization for the introductory example

interesting to see whether the lower complexity obtained in our setting is preserved when going from $\mathcal{EL}$ concepts to $\mathcal{ELI}$ concepts or to acyclic conjunctive queries.

In this paper we have restricted the attention to singleton policies, i.e., ones consisting of a single concept. With such a policy, Ben can for instance prevent people from finding out who are the famous comedians, using the policy concept *Comedian* $\sqcap$ *Famous*. But he cannot prevent them from finding out who is famous or a comedian, since this would require using the non-singleton policy $\{\mathit{Comedian}, \mathit{Famous}\}$. It is currently not clear whether and how our results can be extended from singleton policies to general ones consisting of a finite set of $\mathcal{EL}$ concepts. The papers [2, 5] investigate both compliance and safety for such general policies, but they restrict the data to $\mathcal{EL}$ instance stores. The work in [3] considers general quantified ABoxes and policies, but presents results for compliance only. It would be interesting to find out whether the NP upper bound for deciding safety in this general cases has a matching NP lower bound, and whether our approach for computing optimal safe anonymizations can be extended to this setting. Given a non-singleton policy $\{P_1, \ldots, P_k\}$ and a quantified ABox $\exists X.\mathcal{A}$, one could, of course, first apply our method for computing an optimal safe anonymization for the case of singleton policies to $\exists X.\mathcal{A}$ and $\{P_1\}$, then to the resulting quantified ABox and $\{P_2\}$, etc. While this would indeed yield a quantified ABox that is safe for $\{P_1, \ldots, P_k\}$, this ABox need not be optimal.

References

[1] F. Baader, I. Horrocks, C. Lutz, and U. Sattler. *An Introduction to Description Logic.* Cambridge University Press, 2017.

[2] F. Baader, F. Kriegel, and A. Nuradiansyah. Privacy-preserving ontology publishing for *EL* instance stores. In F. Calimeri, N. Leone, and M. Manna, editors, *Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Proceedings*, volume 11468 of *Lecture Notes in Computer Science*, pages 323–338. Springer, 2019.

[3] F. Baader, F. Kriegel, A. Nuradiansyah, and R. Peñaloza. Computing compliant anonymisations of quantified aboxes w.r.t. $\mathcal{EL}$ policies. In J. Z. P. et al., editor, *The Semantic Web - ISWC 2020 - 19th Int. Semantic Web Conference, 2020, Proceedings, Part I*, volume 12506 of *Lecture Notes in Computer Science*, pages 3–20. Springer, 2020.

[4] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI'97)*, pages 96–101. Morgan Kaufmann, 1999.

[5] F. Baader and A. Nuradiansyah. Mixing description logics in privacy-preserving ontology publishing. In C. Benzmüller and H. Stuckenschmidt, editors, *KI 2019: Advances in Artificial Intelligence - 42nd German Conference on AI*, volume 11793 of *Lecture Notes in Computer Science*, pages 87–100. Springer, 2019.

[6] M. Benedikt, B. C. Grau, and E. V. Kostylev. Source information disclosure in ontology-based data integration. In S. P. Singh and S. Markovitch, editors,

Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pages 1056–1062. AAAI Press, 2017.

[7] P. A. Bonatti and L. Sauro. A confidentiality model for ontologies. In H. A. et al., editor, *The Semantic Web - ISWC 2013 - 12th Int. Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, volume 8218 of *Lecture Notes in Computer Science*, pages 17–32. Springer, 2013.

[8] G. Cima, D. Lembo, R. Rosati, and D. F. Savo. Controlled query evaluation in description logics through instance indistinguishability. In C. Bessiere, editor, *Proceedings of the Twenty-Ninth Int. Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1791–1797. ijcai.org, 2020.

[9] R. Delanaux, A. Bonifati, M. Rousset, and R. Thion. RDF graph anonymization robust to data linkage. In R. Cheng, N. Mamoulis, Y. Sun, and X. Huang, editors, *Web Information Systems Engineering - WISE 2019 - 20th Int. Conference, Proceedings*, volume 11881 of *Lecture Notes in Computer Science*, pages 491–506. Springer, 2019.

[10] C. Dwork. Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, editors, *33rd Int. Colloquium on Automata, Languages and Programming (ICALP 2006)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[11] 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.

[12] B. C. Grau and E. V. Kostylev. Logical foundations of privacy-preserving publishing of linked data. In D. Schuurmans and M. P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 943–949. AAAI Press, 2016.

[13] B. C. Grau and E. V. Kostylev. Logical foundations of linked data anonymisation. *J. Artif. Intell. Res.*, 64:253–314, 2019.

[14] R. Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *Lecture Notes in Computer Science*. Springer, 2001.

[15] L. Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002.