Safety of Quantified ABoxes w.r.t. Singleton $\mathcal{EL}$ Policies

Franz Baader[1] Francesco Kriegel[1] **Adrian Nuradiansyah**[1]
Rafael Peñaloza[2]

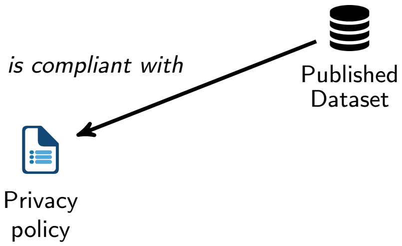
[1]Technische Universität Dresden & [2]University of Milano-Bicocca

In the 36th ACM/SIGAPP Symposium On Applied Computing

March 23rd, 2021

Funded by Deutsche Forschungsgemeinschaft German Research Foundation

An Illustration of Non-Safety

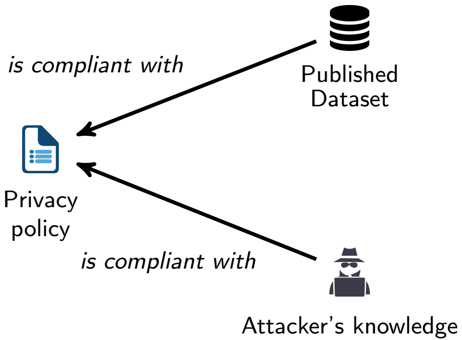


Dataset:
$\exists\{x\}.\{father(BEN, x), Comedian(x)\}$

Policy:
$Comedian \sqcap \exists father.Comedian$

BEN is not an instance of the policy concept w.r.t. the dataset

An Illustration of Non-Safety



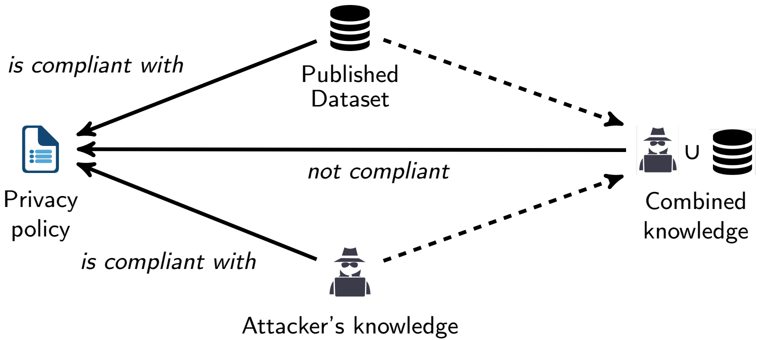
Dataset:
$\exists\{x\}.\{father(BEN, x), Comedian(x)\}$

Policy:
$Comedian \sqcap \exists father.Comedian$

Attacker knows
$\exists\{x\}.\{Comedian(BEN)\}$

BEN is not an instance of the policy concept w.r.t. the attacker's knowledge

An Illustration of Non-Safety



Dataset:
$\exists\{x\}.\{father(BEN, x), Comedian(x)\}$

Policy:
$Comedian \sqcap \exists father.Comedian$

Attacker knows
$\exists\{x\}.\{Comedian(BEN)\}$

BEN is an instance of the policy concept w.r.t. the dataset and the attacker's knowledge $\Rightarrow$ *the dataset is **compliant with**, but **not safe** for the policy !*

What We Want To Do

Our Research Questions

1. How to **decide if a dataset is safe for a policy** i.e.,

   none of the secret information is revealed, even if the attacker has additional compliant knowledge ?

2. How to **anonymise a dataset** such that
   - the anonymised dataset is safe for a policy,
   - all the anonymized information follows from the original dataset, and
   - the amount of lost entailments due to the anonymisation is minimal?

Assumption: Our problems are considered in the context of Description Logics

How our Dataset Looks Like

Our dataset is a **quantified ABox** $\exists X.\mathcal{A}$

Example: $\exists\{x\}.\{Comedian(BEN), father(BEN, x), Comedian(x)\}$

Our dataset is a **quantified ABox** $\exists X.\mathcal{A}$

Example: $\exists\{x\}.\{Comedian(BEN), father(BEN, x), Comedian(x)\}$

$\exists X.\mathcal{A}$ *is built over*

variable names, e.g., $x, y, z, \ldots$

concept names, e.g., *Comedian*, *Actor*, ...

individual names, e.g., *BEN*, *JERRY*, ...

role names, e.g., *mother*, *father*, ...

and the **matrix** $\mathcal{A}$ of the quantified ABox consists of:

- **concept assertions**, e.g., $Comedian(BEN), Actor(x)\ldots$
- **role assertions**, e.g., $mother(BEN, x), father(BEN, y)\ldots$

How our Dataset Looks Like

Our dataset is a **quantified ABox** $\exists X.\mathcal{A}$

Example: $\exists\{x\}.\{\mathit{Comedian}(\mathit{BEN}), \mathit{father}(\mathit{BEN}, x), \mathit{Comedian}(x)\}$

Note:

- Every variable or individual occurring in $\exists X.\mathcal{A}$ is called an **object**
- $\exists X.\mathcal{A} \models \exists Y.\mathcal{B}$ denotes that $\exists X.\mathcal{A}$ **entails** $\exists Y.\mathcal{B}$
- A quantified ABox without variables is a *traditional DL ABox*

How our Policies Look Like

A **policy** P is a **concept of the description logic** $\mathcal{EL}$

Example: $P = \mathit{Comedian} \sqcap \exists \mathit{father}.(\mathit{Comedian} \sqcap \mathit{Actor})$

$$\text{Atoms}(P) = \{\mathit{Comedian}, \exists \mathit{father}.(\mathit{Comedian} \sqcap \mathit{Actor})\}$$

(***concept names*** *or* ***existential restrictions*** *occurring in* P)

Instance Relationships in $\mathcal{EL}$

- $\exists X.\mathcal{A} \models D(u)$ means that the object u is an **instance of** the $\mathcal{EL}$ concept D w.r.t. $\exists X.\mathcal{A}$
- **Instance relationships** in $\mathcal{EL}$ can be checked in polynomial time

A Formal Definition of Safety

In **(Baader, Kriegel, Nuradiansyah, Penaloza, ISWC 2020)**, the notion of **policy-compliance** for quantified ABoxes was introduced

Compliance and Safety

A quantified ABox $\exists X.\mathcal{A}$ is

- **compliant with** a policy concept P iff $\exists X.\mathcal{A} \not\models P(a)$ for all individuals a

A Formal Definition of Safety

In **(Baader, Kriegel, Nuradiansyah, Penaloza, ISWC 2020)**, the notion of **policy-compliance** for quantified ABoxes was introduced

Compliance and Safety

A quantified ABox $\exists X.\mathcal{A}$ is

- **compliant with** a policy concept P iff $\exists X.\mathcal{A} \not\models P(a)$ for all individuals a
- **safe for** P iff for each quantified ABox $\exists Y.\mathcal{B}$ that is compliant with P,

  the union $\exists X.\mathcal{A} \cup \exists Y.\mathcal{B}$ is also compliant with P

What Makes a Quantified ABox Not Safe for a Policy

- **Observation 1**
  There exist an individual a and $B \in \text{Atoms}(P)$ such that $B(a)$ is in $\mathcal{A}$, e.g.,

$$\exists X.\mathcal{A} := \exists\emptyset.\{C(BEN), f(BEN, JERRY)\} \qquad P := C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 2**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a,u) \in \mathcal{A}$ such that u is an individual, e.g.,

$$\exists X.\mathcal{A} = \exists\emptyset.\{f(BEN, JERRY)\} \qquad P = C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN), C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 3**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a,u) \in \mathcal{A}$ such that "a part of D can be homomorphically mapped to $\mathcal{A}$ at u", e.g.,

$$\exists X.\mathcal{A} = \exists\{x\}.\{f(BEN, x), C(x)\} \qquad P = C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN)\} \text{ (an attacker's knowledge)}$$

What Makes a Quantified ABox Not Safe for a Policy

- **Observation 1**
  There exist an individual a and $B \in \text{Atoms}(P)$ such that $B(a)$ is in $\mathcal{A}$, e.g.,

  $$\exists X.\mathcal{A} := \exists\emptyset.\{C(BEN), f(BEN, JERRY)\} \qquad P := C \sqcap \exists f.C$$

  $$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 2**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a, u) \in \mathcal{A}$ such that u is an individual, e.g.,

  $$\exists X.\mathcal{A} = \exists\emptyset.\{f(BEN, JERRY)\} \qquad P = C \sqcap \exists f.C$$

  $$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN), C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 3**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a, u) \in \mathcal{A}$ such that "a part of D can be homomorphically mapped to $\mathcal{A}$ at u", e.g.,

  $$\exists X.\mathcal{A} = \exists\{x\}.\{f(BEN, x), C(x)\} \qquad P = C \sqcap \exists f.C$$

  $$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN)\} \text{ (an attacker's knowledge)}$$

What Makes a Quantified ABox Not Safe for a Policy

- **Observation 1**
  There exist an individual a and $B \in \text{Atoms}(P)$ such that $B(a)$ is in $\mathcal{A}$, e.g.,

$$\exists X.\mathcal{A} := \exists\emptyset.\{C(BEN), f(BEN, JERRY)\} \qquad P := C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 2**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a, u) \in \mathcal{A}$ such that u is an individual, e.g.,

$$\exists X.\mathcal{A} = \exists\emptyset.\{f(BEN, JERRY)\} \qquad P = C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN), C(JERRY)\} \text{ (an attacker's knowledge)}$$

- **Observation 3**
  There exist an individual a, an atom $\exists r.D \in \text{Atoms}(P)$, and $r(a, u) \in \mathcal{A}$ such that "a part of D can be homomorphically mapped to $\mathcal{A}$ at u", e.g.,

$$\exists X.\mathcal{A} = \exists\{x\}.\{f(BEN, x), C(x)\} \qquad P = C \sqcap \exists f.C$$

$$\exists X'.\mathcal{A}' := \exists\emptyset.\{C(BEN)\} \text{ (an attacker's knowledge)}$$

Partial Homomorphism

- **Observation 2**
  There exist an individual a, an atom $\exists r.D \in \mathrm{Atoms}(P)$, *and* $r(a,u) \in \mathcal{A}$ *such that* u is an individual e.g.,
- **Observation 3**
  There exist an individual a, an atom $\exists r.D \in \mathrm{Atoms}(P)$, *and* $r(a,u) \in \mathcal{A}$ *such that* "a part of D can be homomorphically mapped to $\mathcal{A}$ at u"

The two conditions above formally are called **the existence of a partial homomorphism from D to $\exists X.\mathcal{A}$ at u**

Partial Homomorphism

- **Observation 2**
  There exist an individual a*, an atom* $\exists r.D \in \mathrm{Atoms}(P)$*, and* $r(a,u) \in \mathcal{A}$ *such that* u is an individual e.g.,
- **Observation 3**
  There exist an individual a*, an atom* $\exists r.D \in \mathrm{Atoms}(P)$*, and* $r(a,u) \in \mathcal{A}$ *such that* "a part of D can be homomorphically mapped to $\mathcal{A}$ at u"

The two conditions above formally are called **the existence of a partial homomorphism from D to $\exists X.\mathcal{A}$ at u**

The Existence of a Partial Homomorphism

Checking the **existence of a partial homomorphism** can be done in **polynomial time**

Deciding if an ABox is safe for a policy

Characterizing Safety

$\exists X.\mathcal{A}$ is safe for a policy P iff for each individual name a

1. if $B \in \text{Atoms}(P)$, then **the assertion** $B(a)$ **is not in** $\mathcal{A}$
2. if role assertion $r(a, u) \in \mathcal{A}$ and $\exists r.D \in \text{Atoms}(P)$, then there is **no partial homomorphism** from D to $\exists X.\mathcal{A}$ at u.

Complexity of the Safety Problem

Checking if a quantified ABox is safe for a policy concept can be done in **polynomial time**

Optimal Safe Anonymisations

The ABox

$$\exists\{x\}.\{father(BEN, x)\}$$

is safe for the policy $Comedian \sqcap \exists father.Comedian$. However, the following ABox

$$\exists\{x, y\}.\{father(BEN, x), Comedian(y), father(y, x)\}$$

is also safe for the policy and **entails the first ABox.**

Optimal Safe Anonymisations

The ABox

$$\exists\{x\}.\{father(BEN, x)\}$$

is safe for the policy $Comedian \sqcap \exists father.Comedian$. However, the following ABox

$$\exists\{x, y\}.\{father(BEN, x), Comedian(y), father(y, x)\}$$

is also safe for the policy and **entails the first ABox.**

A **quantified ABox** $\exists Y.\mathcal{B}$ is an **optimal safe anonymisation** of $\exists X.\mathcal{A}$ for a policy P iff

- $\exists Y.\mathcal{B}$ is safe for P (**safety**)
- $\exists X.\mathcal{A} \models \exists Y.\mathcal{B}$ (**anonymisation**)
- there is no safe anonymisation $\exists Z.\mathcal{C}$ of $\exists X.\mathcal{A}$ for P that strictly entails $\exists Y.\mathcal{B}$ (**optimality**)

Computing an Optimal Safe Anonymisation

$$\exists X.\mathcal{A} := \exists\emptyset.\{\mathit{Comedian}(\mathit{BEN}), \mathit{father}(\mathit{BEN}, \mathit{JERRY}), \mathit{Comedian}(\mathit{JERRY})\}$$
$$P := \mathit{Comedian} \sqcap \exists \mathit{father}.\mathit{Comedian}$$

Computing an Optimal Safe Anonymisation

$$\exists X.\mathcal{A} := \exists\emptyset.\{\mathit{Comedian}(\mathit{BEN}), \mathit{father}(\mathit{BEN}, \mathit{JERRY}), \mathit{Comedian}(\mathit{JERRY})\}$$
$$P := \mathit{Comedian} \sqcap \exists \mathit{father}.\mathit{Comedian}$$

The main idea of the approach:

1.) For each object u in $\exists X.\mathcal{A}$, **introduce copies** $y_{u,\mathcal{K}}$ **of them** as a variable in $\exists Y.\mathcal{B}$, where $\mathcal{K} \subseteq \mathsf{Atoms}(P)$

it is sufficient to create at most **exponentially many such copies**

Computing an Optimal Safe Anonymisation

$$\exists X.\mathcal{A} := \exists\emptyset.\{Comedian(BEN), father(BEN, JERRY), Comedian(JERRY)\}$$
$$P := Comedian \sqcap \exists father.Comedian$$

b $\quad y_{b,\emptyset} \quad y_{b,\{C\}} \quad y_{b,\{\exists f.C\}} \quad y_{b,\{C,\exists f.C\}}$

j $\quad y_{j,\emptyset} \quad y_{j,\{C\}} \quad y_{j,\{\exists f.C\}} \quad y_{j,\{C,\exists f.C\}}$

Computing an Optimal Safe Anonymisation

$$\exists X.\mathcal{A} := \exists\emptyset.\{Comedian(BEN), father(BEN, JERRY), Comedian(JERRY)\}$$
$$P := Comedian \sqcap \exists father.Comedian$$

The main idea of the approach:

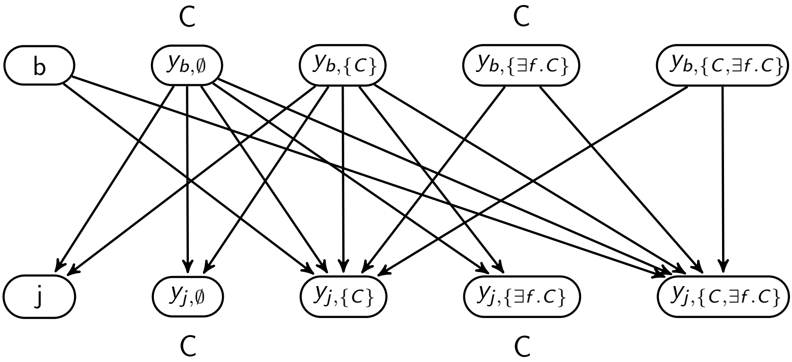
2.) For each individual a, b and each variable $y_{u,\mathcal{K}}$ in $\exists Y.\mathcal{B}$, **ensure that they satisfy less assertions**, in particular

- if $B(a)$ in $\exists X.\mathcal{A}$ and $B \in \text{Atoms}(P)$, then don't add $B(a)$ in $\exists Y.\mathcal{B}$
- if $r(a, b)$ in $\exists X.\mathcal{A}$ and $\exists r.D \in \text{Atoms}(P)$, then don't add $r(a, b)$ in $\exists Y.\mathcal{B}$ and
- if $D \in \mathcal{K}$, then no partial homomorphism from D to $\exists Y.\mathcal{B}$ at $y_{u,\mathcal{K}}$

Computing an Optimal Safe Anonymisation

$$\exists X.\mathcal{A} := \exists\emptyset.\{Comedian(BEN), father(BEN, JERRY), Comedian(JERRY)\}$$
$$P := Comedian \sqcap \exists father.Comedian$$

The Optimal Safe Anonymisation $\exists Y.\mathcal{B}$ of $\exists X.\mathcal{A}$ for P



Complexity of Computing The Optimal Safe Anonymisation

Results for the Computational Problem

1. For a quantified ABox $\exists X.\mathcal{A}$ and a policy concept P, the optimal safe anonymisation of $\exists X.\mathcal{A}$ for P is **unique** (up to equivalence)
2. The optimal safe anonymisation can be computed in
   - **exponential time** for **combined complexity**
   - **polynomial time** for **data complexity** i.e., the size of P is fixed

Future Work and References

Future Work:

- Extending the **expressiveness of the policies**
  e.g., $\mathcal{EL} \rightarrow \mathcal{ELI}$, i.e., $\mathcal{EL}$ with inverse roles
- Extending our results to **non-singleton policies**, i.e., policies that have more than one concept
- Adding static **background knowledge (TBoxes)** to both published quantified ABox and the attackers' knowledge

Future Work and References

Future Work:

- Extending the **expressiveness of the policies**
  e.g., $\mathcal{EL} \rightarrow \mathcal{ELI}$, i.e., $\mathcal{EL}$ with inverse roles
- Extending our results to **non-singleton policies**, i.e., policies that have more than one concept
- Adding static **background knowledge (TBoxes)** to both published quantified ABox and the attackers' knowledge

Our work is based on the following **related work**:

- F. Baader, F. Kriegel, A. Nuradiansyah, R. Peñaloza, *Computing Compliant Anonymisations of Quantified ABoxes w.r.t. $\mathcal{EL}$ Policies*, ISWC 2020
- B. Cuenca Grau and E. Kostylev, *Logical Foundations of Linked Data Anonymizations*, JAIR, 2019