Towards Privacy-Preserving Ontology Publishing

F. Baader & **A. Nuradiansyah**

Technische Universität Dresden

October 27, 2018

Privacy-Preserving Ontology Publishing

- In privacy, **repair** may not be enough!
- Given an **ontology** $\mathfrak{O}$, a **policy** $\mathcal{P} = \{\alpha_1, \ldots, \alpha_n\}$ is a finite set of axioms to be hidden, i.e., an attacker **should not be able to see** α_i as a consequence of $\mathfrak{O}$.

Privacy-Preserving Ontology Publishing

- In privacy, **repair** may not be enough!
- Given an **ontology** $\mathfrak{O}$, a **policy** $\mathcal{P} = \{\alpha_1, \ldots, \alpha_n\}$ is a finite set of axioms to be hidden, i.e., an attacker **should not be able to see** α_i as a consequence of $\mathfrak{O}$.
- Suppose $\mathfrak{O} \models \alpha_i$ for some $\alpha_i \in \mathcal{P}$ i.e., $\mathfrak{O}$ **does not comply with** $\mathcal{P}$.
- Let $\mathfrak{O}'$ be a **repair** of $\mathfrak{O}$ w.r.t. α_i such that $\mathfrak{O}' \not\models \alpha_i$ for all i.

Privacy-Preserving Ontology Publishing

- In privacy, **repair** may not be enough!
- Given an **ontology** $\mathfrak{O}$, a **policy** $\mathcal{P} = \{\alpha_1, \ldots, \alpha_n\}$ is a finite set of axioms to be hidden, i.e., an attacker **should not be able to see** α_i as a consequence of $\mathfrak{O}$.
- Suppose $\mathfrak{O} \models \alpha_i$ for some $\alpha_i \in \mathcal{P}$ i.e., $\mathfrak{O}$ **does not comply with** $\mathcal{P}$.
- Let $\mathfrak{O}'$ be a **repair** of $\mathfrak{O}$ w.r.t. α_i such that $\mathfrak{O}' \not\models \alpha_i$ for all i.
- But, when $\mathfrak{O}'$ is **published** on the Web, ...
  an attacker may know an ontology $\mathfrak{O}''$ such that $\mathfrak{O}'' \not\models \alpha_i$, but $\mathfrak{O}' \cup \mathfrak{O}'' \models \alpha_i$.
- In this case, it is still not safe to publish $\mathfrak{O}'$.

Privacy-Preserving Ontology Publishing

What people already did:

In **(Cuenca Grau & Kostylev, 2016):**

- Privacy-Preserving Data Publishing
- Information to be published: a relational dataset with (labeled) nulls
- Policy is a conjunctive query.
- Considering three privacy properties when publishing datasets: **policy-compliant, policy-safety, and optimality**.
- Published information does not have background knowledge.

What we want to do:

- **Privacy-Preserving Ontology Publishing (PPOP)**
- Addressed in the context of **Description Logic Ontologies**

PPOP with Role-Free ABoxes in $\mathcal{EL}$

- **Starting point**: $\mathcal{EL}$ Ontologies with **role-free ABoxes** and empty TBoxes.
- An ABox $\mathcal{A}$ is **role-free** if all the axioms $\beta \in \mathcal{A}$ are only in the form of $D(a)$.
- W.l.o.g., only **one concept assertion** in $\mathcal{A}$ speaks about one individual

$$\text{If } C_1(a) \in \mathcal{A} \text{ and } C_2(a) \in \mathcal{A}, \text{ then } (C_1 \sqcap C_2)(a) \in \mathcal{A}$$

- Safe Ontologies $\xrightarrow{reduced}$ Safe Concepts
- Information to be published for an individual a: an $\mathcal{EL}$ concept C
- **Policy** is a finite set of $\mathcal{EL}$ concepts $D_1, \ldots, D_p$, such that $D_i \not\equiv \top$ for all $i \in \{1, \ldots, p\}$.

Compliance, Safety, and Optimality

Given a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$ and an $\mathcal{EL}$ concept C, the $\mathcal{EL}$ concept C' is

- **compliant** with $\mathcal{P}$ if $C' \not\sqsubseteq D_i$ for all $i \in \{1, \ldots, p\}$.
- **safe** for $\mathcal{P}$ if $C' \sqcap C''$ is compliant with $\mathcal{P}$ for all $\mathcal{EL}$-concepts C'' that are compliant with $\mathcal{P}$.

Compliance, Safety, and Optimality

Given a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$ and an $\mathcal{EL}$ concept C, the $\mathcal{EL}$ concept C' is

- **compliant** with $\mathcal{P}$ if $C' \not\sqsubseteq D_i$ for all $i \in \{1, \ldots, p\}$.
- **safe** for $\mathcal{P}$ if $C' \sqcap C''$ is compliant with $\mathcal{P}$ for all $\mathcal{EL}$-concepts C'' that are compliant with $\mathcal{P}$.
- a **$\mathcal{P}$-compliant (safe) generalization** of C if
  - $C \sqsubseteq C'$ and
  - C' is compliant with (safe for) $\mathcal{P}$.

Compliance, Safety, and Optimality

Given a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$ and an $\mathcal{EL}$ concept C, the $\mathcal{EL}$ concept C' is

- **compliant** with $\mathcal{P}$ if $C' \not\sqsubseteq D_i$ for all $i \in \{1, \ldots, p\}$.
- **safe** for $\mathcal{P}$ if $C' \sqcap C''$ is compliant with $\mathcal{P}$ for all $\mathcal{EL}$-concepts C'' that are compliant with $\mathcal{P}$.
- a **$\mathcal{P}$-compliant (safe) generalization** of C if
  - $C \sqsubseteq C'$ and
  - C' is compliant with (safe for) $\mathcal{P}$.
- a **$\mathcal{P}$-optimal compliant (safe) generalization** of C if
  - $C \sqsubseteq C'$,
  - C' is a $\mathcal{P}$-compliant (safe) generalization of C, and
  - there is no $\mathcal{P}$-compliant (safe) generalization of C s.t. $C'' \sqsubset C'$.

Illustration on Compliance, Safety, and Optimality

- Consider a **policy** $\mathcal{P} = \{D\}$ specifying what information should be kept "secret" about *linda*

$$D = Patient \sqcap \exists seen\_by.(Doctor \sqcap \exists works\_in.Cardiology)$$

- Assume information C is published about *linda*

$$C = Patient \sqcap Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.Cardiology)$$

Note C is not compliant with D, i.e., $C \sqsubseteq D$.

Illustration on Compliance, Safety, and Optimality

- Consider a **policy** $\mathcal{P} = \{D\}$ specifying what information should be kept "secret" about *linda*

$$D = Patient \sqcap \exists seen\_by.(Doctor \sqcap \exists works\_in.Cardiology)$$

- Assume information C is published about *linda*

$$C = Patient \sqcap Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.Cardiology)$$

  Note C is not compliant with D, i.e., $C \sqsubseteq D$.

- Generalizing C to C_1 yields a compliant concept

$$C_1 = Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.Cardiology)$$

  But, C_1 is **not safe for** D since if the attacker knows $Patient(linda)$, then $C_1 \sqcap Patient \sqsubseteq D$ is revealed.

Illustration on Compliance, Safety, and Optimality

- Consider a **policy** $\mathcal{P} = \{D\}$ specifying what information should be kept "secret" about *linda*

$$D = Patient \sqcap \exists seen\_by.(Doctor \sqcap \exists works\_in.Cardiology)$$

- Assume information C is published about *linda*

$$C = Patient \sqcap Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.Cardiology)$$

  Note C is not compliant with D, i.e., $C \sqsubseteq D$.

- Let us **make it safe**!

$$C_2 = Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.\top)$$

  But, C_2 is still not optimal since more information than necessary is removed.

Illustration on Compliance, Safety, and Optimality

- Consider a **policy** $\mathcal{P} = \{D\}$ specifying what information should be kept "secret" about *linda*

$$D = Patient \sqcap \exists seen\_by.(Doctor \sqcap \exists works\_in.Cardiology)$$

- Assume information C is published about *linda*

$$C = Patient \sqcap Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.Cardiology)$$

  Note C is not compliant with D, i.e., $C \sqsubseteq D$.

- Let us **make it safe**!

$$C_2 = Female \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.\top)$$

  But, C_2 is still not optimal since more information than necessary is removed.

- Make it **optimal**!

$$\begin{aligned} C_3 = Female \quad & \sqcap \exists seen\_by.(Doctor \sqcap Male \sqcap \exists works\_in.\top) \\ & \sqcap \exists seen\_by.(Male \sqcap \exists works\_in.Cardiology) \end{aligned}$$

Characterizing Compliant

- Let $\mathtt{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.

Characterizing Compliant

- Let $\texttt{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.
- $\texttt{con}(C)$ covers $\texttt{con}(D)$ iff for all $F \in \texttt{con}(D)$, there is $E \in \texttt{con}(C)$ such that $E \sqsubseteq F$

Characterizing Compliant

- Let $\texttt{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.
- $\texttt{con}(C)$ covers $\texttt{con}(D)$ iff for all $F \in \texttt{con}(D)$, there is $E \in \texttt{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

Characterizing Compliant

- Let $\texttt{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.
- $\texttt{con}(C)$ covers $\texttt{con}(D)$ iff for all $F \in \texttt{con}(D)$, there is $E \in \texttt{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

Compliance

C is **compliant** with $\mathcal{P}$ iff $\texttt{con}(C)$ does not cover $\texttt{con}(D_i)$ for any $i \in \{1, \ldots, p\}$.

Characterizing Compliant

- Let $\texttt{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.
- $\texttt{con}(C)$ covers $\texttt{con}(D)$ iff for all $F \in \texttt{con}(D)$, there is $E \in \texttt{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

Compliance

C is **compliant** with $\mathcal{P}$ iff $\texttt{con}(C)$ does not cover $\texttt{con}(D_i)$ for any $i \in \{1, \ldots, p\}$.

Complexity for Compliance

- Deciding whether C' is compliant w.r.t. $\mathcal{P}$ is in **PTime.**

Characterizing Compliant

- Let $\mathrm{con}(C)$ be the set of all **atoms** A or $\exists r.E$ occurring in the **top-level conjunction** of C.
- $\mathrm{con}(C)$ covers $\mathrm{con}(D)$ iff for all $F \in \mathrm{con}(D)$, there is $E \in \mathrm{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

Compliance

C is **compliant** with $\mathcal{P}$ iff $\mathrm{con}(C)$ does not cover $\mathrm{con}(D_i)$ for any $i \in \{1, \ldots, p\}$.

Complexity for Compliance

- Deciding whether C' is compliant w.r.t. $\mathcal{P}$ is in **PTime.**
- One optimal $\mathcal{P}$-compliant generalization can be **computed in ExpTime**.
- The set of all optimal $\mathcal{P}$-compliant generalizations can be **computed in ExpTime**.

Characterizing Safety

Assume $\mathcal{P}$ is **redundant-free**: every $D_i, D_j \in \mathcal{P}$ are **incomparable w.r.t. subsumption**.

Characterizing Safety

Assume $\mathcal{P}$ is **redundant-free**: every $D_i, D_j \in \mathcal{P}$ are **incomparable w.r.t. subsumption**.

Safety

C' is safe for $\mathcal{P}$ iff there is **no pair of atoms** (E, F) such that

$$E \in \mathtt{con}(C'),\ F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ and } E \sqsubseteq F$$

Deciding whether C' is safe for $\mathcal{P}$ is in **PTime.**

Characterizing Safety

Assume $\mathcal{P}$ is **redundant-free**: every $D_i, D_j \in \mathcal{P}$ are **incomparable w.r.t. subsumption**.

Safety

C' is safe for $\mathcal{P}$ iff there is **no pair of atoms** (E, F) such that

$$E \in \mathtt{con}(C'),\ F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ and } E \sqsubseteq F$$

Deciding whether C' is safe for $\mathcal{P}$ is in **PTime.**

The Optimal $\mathcal{P}$-Safe Generalization

- If C'_1, C'_2 are $\mathcal{P}$-safe generalizations of C, then $C'_1 \sqcap C'_2$ is also a $\mathcal{P}$-safe generalization of C.
  $\Rightarrow$ Optimal $\mathcal{P}$-safe generalization is **unique up to equivalence**.

Characterizing Safety

Assume $\mathcal{P}$ is **redundant-free**: every $D_i, D_j \in \mathcal{P}$ are **incomparable w.r.t. subsumption**.

Safety

C' is safe for $\mathcal{P}$ iff there is **no pair of atoms** (E, F) such that

$$E \in \mathtt{con}(C'),\ F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ and } E \sqsubseteq F$$

Deciding whether C' is safe for $\mathcal{P}$ is in **PTime.**

The Optimal $\mathcal{P}$-Safe Generalization

- If C_1', C_2' are $\mathcal{P}$-safe generalizations of C, then $C_1' \sqcap C_2'$ is also a $\mathcal{P}$-safe generalization of C.
  $\Rightarrow$ Optimal $\mathcal{P}$-safe generalization is **unique up to equivalence**.
- The $\mathcal{P}$-optimal safe generalization of C can be **computed in ExpTime.**
  $\Rightarrow$ Requiring the computation of optimal $\mathcal{P}$-compliant generalizations.

Future Work

- Decision problem for optimality
- Considering PPOP with $\mathcal{EL}$ concepts w.r.t. (Acylic) TBoxes
- Considering a setting where $\mathcal{A}$ contains concept and role assertions
- Considering $\mathcal{ELO}$ concepts

Thank You

ROSI