Towards Privacy-Preserving Ontology Publishing

F. Baader & A. Nuradiansyah

Technische Universität Dresden

October 27, 2018
In privacy, **repair** may not be enough!

Given an **ontology** $\mathcal{O}$, 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 $\alpha_i$ as a consequence of $\mathcal{O}$. 

But, when $\mathcal{O}'$ is published on the Web, ...
In privacy, **repair** may not be enough!

Given an **ontology** $\mathcal{O}$, 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 $\alpha_i$ as a consequence of $\mathcal{O}$.

Suppose $\mathcal{O} \models \alpha_i$ for some $\alpha_i \in \mathcal{P}$ i.e., $\mathcal{O}$ does not comply with $\mathcal{P}$.

Let $\mathcal{O}'$ be a **repair** of $\mathcal{O}$ w.r.t. $\alpha_i$ such that $\mathcal{O}' \not\models \alpha_i$ for all $i$. 
In privacy, **repair** may not be enough!

Given an **ontology** $\mathcal{O}$, 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** $\alpha_i$ as a consequence of $\mathcal{O}$.

Suppose $\mathcal{O} \models \alpha_i$ for some $\alpha_i \in \mathcal{P}$ i.e., $\mathcal{O}$ **does not comply** with $\mathcal{P}$.

Let $\mathcal{O}'$ be a **repair** of $\mathcal{O}$ w.r.t. $\alpha_i$ such that $\mathcal{O}' \not\models \alpha_i$ for all $i$.

But, when $\mathcal{O}'$ is **published** on the Web, ... an attacker may know an ontology $\mathcal{O}''$ such that $\mathcal{O}'' \not\models \alpha_i$, but $\mathcal{O}' \cup \mathcal{O}'' \models \alpha_i$.

In this case, it is still not safe to publish $\mathcal{O}'$. 
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

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

- Safe Ontologies $\xrightarrow{\text{reduced}}$ 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 $i \in \{1, \ldots, p\}$. 
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 $\mathcal{P}$.
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 (safe) for $\mathcal{P}$. 
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 $C' \not\subseteq D_i$ 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
  - $C \subseteq C'$ and
  - $C'$ is compliant with (safe for) $\mathcal{P}$.

- a $\mathcal{P}$-optimal compliant (safe) generalization of $C$ if
  - $C \subseteq C'$,
  - $C'$ is a $\mathcal{P}$-compliant (safe) generalization of $C$, and
  - there is no $\mathcal{P}$-compliant (safe) generalization of $C$ s.t. $C'' \sqsubset C'$.
Consider a policy $\mathcal{P} = \{D\}$ specifying what information should be kept “secret” about Linda

$$D = Patient \sqcap \exists \text{seen by.}(Doctor \sqcap \exists \text{works in. Cardiology})$$

Assume information $C$ is published about Linda

$$C = Patient \sqcap Female \sqcap \exists \text{seen by.}(Doctor \sqcap Male \sqcap \exists \text{works in. Cardiology})$$

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

Illustration on Compliance, Safety, and Optimality

- Consider a policy \( \mathcal{P} = \{D\} \) specifying what information should be kept “secret” about Linda

\[
D = \text{Patient} \sqcap \exists \text{seen\_by}.(\text{Doctor} \sqcap \exists \text{works\_in.Cardiology})
\]

- Assume information \( C \) is published about Linda

\[
C = \text{Patient} \sqcap \text{Female} \sqcap \exists \text{seen\_by}.(\text{Doctor} \sqcap \text{Male} \sqcap \exists \text{works\_in.Cardiology})
\]
Note \( C \) is not compliant with \( D \), i.e., \( C \sqsubseteq D \).

- Generalizing \( C \) to \( C_1 \) yields a compliant concept

\[
C_1 = \text{Female} \sqcap \exists \text{seen\_by}.(\text{Doctor} \sqcap \text{Male} \sqcap \exists \text{works\_in.Cardiology})
\]
But, \( C_1 \) is not safe for \( D \) since if the attacker knows \( \text{Patient}(linda) \), then \( C_1 \sqcap \text{Patient} \sqsubseteq D \) is revealed.
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., $C \subseteq D$.

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

$$D = Patient \sqcap \exists \text{seen}_\text{by}.(Doctor \sqcap \exists \text{works}_\text{in}.Cardiology)$$

Assume information $C$ is published about *linda*

$$C = Patient \sqcap Female \sqcap \exists \text{seen}_\text{by}.(Doctor \sqcap Male \sqcap \exists \text{works}_\text{in}.Cardiology)$$

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

Let us **make it safe**!

$$C_2 = Female \sqcap \exists \text{seen}_\text{by}.(Doctor \sqcap Male \sqcap \exists \text{works}_\text{in}.\top)$$

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

Make it **optimal**!

$$C_3 = Female \quad \sqcap \exists \text{seen}_\text{by}.(Doctor \sqcap Male \sqcap \exists \text{works}_\text{in}.\top)$$

$$\quad \sqcap \exists \text{seen}_\text{by}.(Male \sqcap \exists \text{works}_\text{in}.Cardiology)$$
Let $\text{con}(C)$ be the set of all atoms $A$ or $\exists r.E$ occurring in the top-level conjunction of $C$. 
Let $\text{con}(C)$ be the set of all atoms $A$ or $\exists r.E$ occurring in the top-level conjunction of $C$.

$\text{con}(C)$ covers $\text{con}(D)$ iff for all $F \in \text{con}(D)$, there is $E \in \text{con}(C)$ such that $E \sqsubseteq F$. 

Compliance
C is compliant with $P$ iff $\text{con}(C)$ does not cover $\text{con}(D)$ for any $i \in \{1, \ldots, p\}$. 

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

F. Baader & A. Nuradiansyah
DL 2018
October 27, 2018
Characterizing Compliant

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

$\text{con}(C)$ covers $\text{con}(D)$ iff for all $F \in \text{con}(D)$, there is $E \in \text{con}(C)$ such that $E \sqsubseteq F \Rightarrow \text{Characterizing } C \sqsubseteq D$.

### Compliance

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

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

- $\text{con}(C)$ covers $\text{con}(D)$ iff for all $F \in \text{con}(D)$, there is $E \in \text{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

Compliance

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

Complexity for Compliance

- Deciding whether $C'$ is compliant w.r.t. $\mathcal{P}$ is in $\text{PTime}$.
Let $\text{con}(C)$ be the set of all atoms $A$ or $\exists r. E$ occurring in the top-level conjunction of $C$.

$\text{con}(C)$ covers $\text{con}(D)$ iff for all $F \in \text{con}(D)$, there is $E \in \text{con}(C)$ such that $E \sqsubseteq F \Rightarrow$ Characterizing $C \sqsubseteq D$.

**Compliance**

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

**Complexity for Compliance**

- Deciding whether $C'$ is compliant w.r.t. $\mathcal{P}$ is in $\text{PTime}$.
- One optimal $\mathcal{P}$-compliant generalization can be computed in $\text{ExpTime}$.
- The set of all optimal $\mathcal{P}$-compliant generalizations can be computed in $\text{ExpTime}$.
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 \textbf{redundant-free}: every $D_i, D_j \in \mathcal{P}$ are \textit{incomparable w.r.t. subsumption}.

\begin{itemize}
  \item \textbf{Safety}
  \begin{itemize}
    \item $\mathcal{C}'$ is safe for $\mathcal{P}$ iff there is \textbf{no pair of atoms} $(E, F)$ such that $E \in \text{con}(\mathcal{C}')$, $F \in \text{con}(D_1) \cup \ldots \cup \text{con}(D_p)$ and $E \sqsubseteq F$
  \end{itemize}
\end{itemize}

Deciding whether $\mathcal{C}'$ is safe for $\mathcal{P}$ is in \textbf{PTime}.
Assume $\mathcal{P}$ is \textbf{redundant-free}: every $D_i, D_j \in \mathcal{P}$ are \textbf{incomparable w.r.t. subsumption}.

\textbf{Safety}

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

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

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

\textbf{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 \textbf{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

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

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. (Acyclic) TBoxes
- Considering a setting where $\mathcal{A}$ contains concept and role assertions
- Considering $\mathcal{ELO}$ concepts
Thank You

ROSI