

Mixing Description Logics in Privacy-Preserving Ontology Publishing

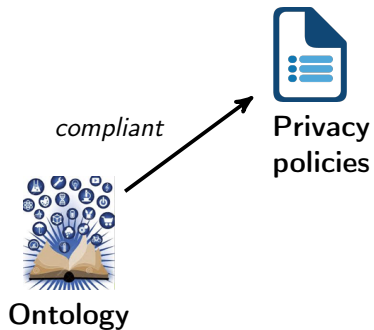
Franz Baader **Adrian Nuradiansyah**

Technische Universität Dresden

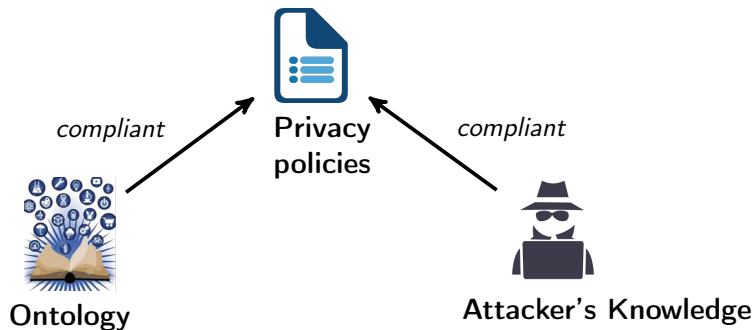
September 25, 2019



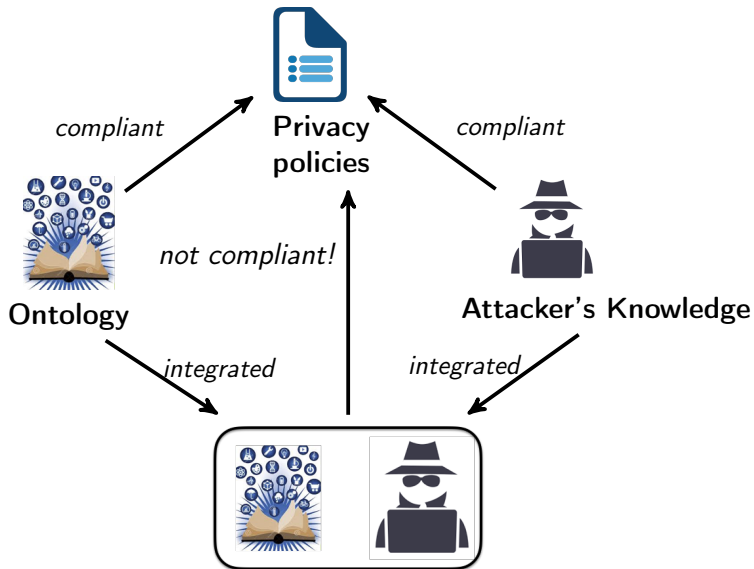
Privacy-Preserving Ontology Publishing (PPOP)



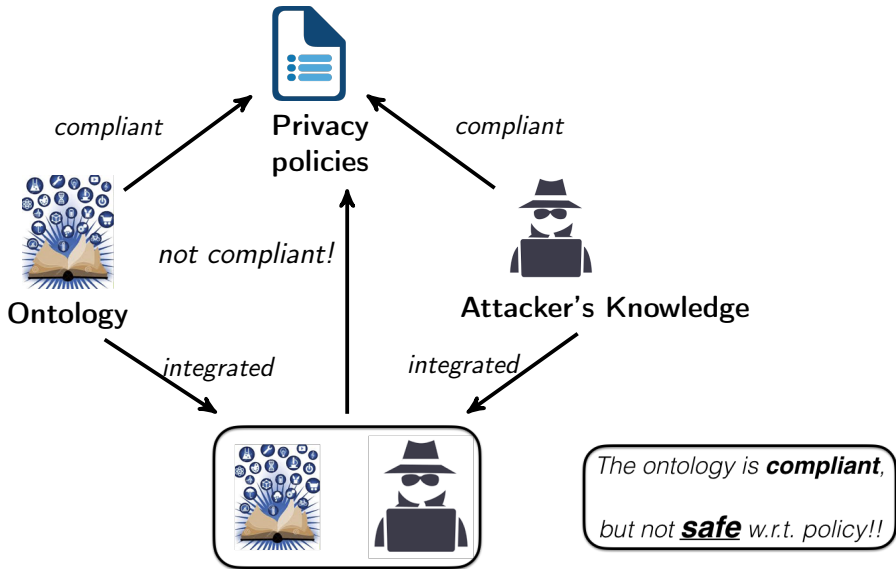
Privacy-Preserving Ontology Publishing (PPOP)



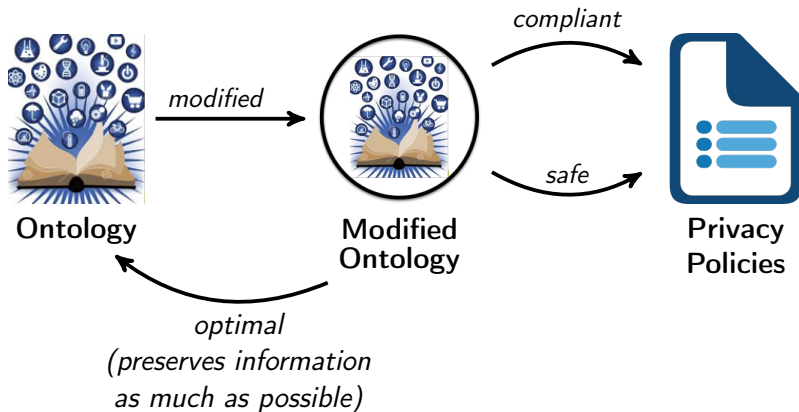
Privacy-Preserving Ontology Publishing (PPOP)



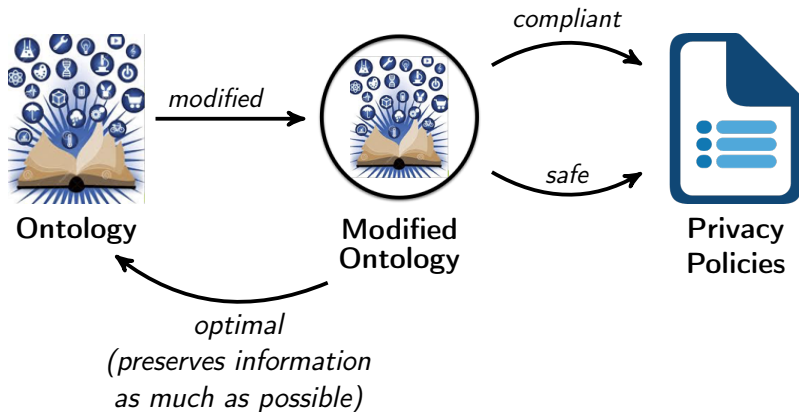
Privacy-Preserving Ontology Publishing (PPOP)



Privacy-Preserving Ontology Publishing (PPOP)



Privacy-Preserving Ontology Publishing (PPOP)



Assumption: Ontologies are formulated in Description Logics (DLs).

What are DLs?

Description Logics

- The **logical underpinning** of **Web Ontology Language (OWL)**
- Commonly used in **medical ontologies**
- Decidable fragments of First Order Logics

- The **logical underpinning** of **Web Ontology Language (OWL)**
- Commonly used in **medical ontologies**
- Decidable fragments of First Order Logics
- The **basic building blocks** are:
 - N_C : set of **concept names** A : *Female, Doctor, Patient, ...*
 - N_R : set of **role names** r : *seenBy, suffer, hasSymptom, ...*
 - N_I : set of **individual names** a : *LINDA, CANCER ...*

- The **logical underpinning** of **Web Ontology Language (OWL)**
- Commonly used in **medical ontologies**
- Decidable fragments of First Order Logics
- The **basic building blocks** are:
 - N_C : set of **concept names** A : *Female, Doctor, Patient, ...*
 - N_R : set of **role names** r : *seenBy, suffer, hasSymptom, ...*
 - N_I : set of **individual names** a : *LINDA, CANCER ...*
- The formal semantics is introduced by means of an **interpretation** ($\mathcal{I} = \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}$)
 - $\Delta^{\mathcal{I}}$: Non-empty domain elements
 - $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
 - $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
 - $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- Using N_C , N_R , and N_I as well as necessary constructors, the notion of **DL concepts** C, D, E are built.

- A DL ontology \mathcal{O} consists of a **TBox** \mathcal{T} and an **ABox** \mathcal{A}
- A TBox \mathcal{T} is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$
→ hierarchical relationship between concepts
- An ABox \mathcal{A} is a set of **concept assertions** $C(a)$ and **role assertions** $r(a, b)$
→ knowledge about individuals

- A DL ontology \mathcal{O} consists of a **TBox** \mathcal{T} and an **ABox** \mathcal{A}
- A TBox \mathcal{T} is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$
→ hierarchical relationship between concepts
- An ABox \mathcal{A} is a set of **concept assertions** $C(a)$ and **role assertions** $r(a, b)$
→ knowledge about individuals
- A DL **Instance Store** \mathcal{O}' is a DL ontology without role assertions

- A DL ontology \mathfrak{D} consists of a **TBox** \mathcal{T} and an **ABox** \mathcal{A}
- A TBox \mathcal{T} is a set of **General Concept Inclusions (GCIs)** $C \sqsubseteq D$
→ hierarchical relationship between concepts
- An ABox \mathcal{A} is a set of **concept assertions** $C(a)$ and **role assertions** $r(a, b)$
→ knowledge about individuals
- A DL **Instance Store** \mathfrak{D}' is a DL ontology without role assertions
- A main reasoning task in DLs \Rightarrow Deciding **subsumption** between concepts
- A concept C is **subsumed** by a concept D , denoted by $C \sqsubseteq D$, iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all interpretations \mathcal{I} .

- $\mathcal{FL}\mathcal{E}$ concepts $C ::= \top$ (*top*) | A | $C \sqcap C$ (*conjunction*) | $\exists r.C$ (**existential restriction**) | $\forall r.C$ (**universal restriction**)
- **Semantics** of some $\mathcal{FL}\mathcal{E}$ concepts:
 - $(\exists r.C)^{\mathcal{I}} = \{d \mid \text{there is } e \in \Delta^{\mathcal{I}} \text{ such that } (d, e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\}$
 - $(\forall r.C)^{\mathcal{I}} = \{d \mid \text{for all } e \in \Delta^{\mathcal{I}} \text{ if } (d, e) \in r^{\mathcal{I}}, \text{ then } e \in C^{\mathcal{I}}\}$
- **Fragments** of $\mathcal{FL}\mathcal{E}$:
 - the DL \mathcal{EL} (excluding value restrictions)
 - the DL \mathcal{FL}_0 (excluding existential restrictions)



\mathcal{EL} Instance Stores
without TBox



\mathcal{EL} Instance Stores
without TBox



$C_1(a), C_2(a)$ implies $(C_1 \sqcap C_2)(a)$

only one concept assertion
speaking about one individual

Problem Setting: PPOP for \mathcal{EL} Instance Stores



\mathcal{EL} Instance Stores
without TBox



$C_1(a), C_2(a)$ implies $(C_1 \sqcap C_2)(a)$

only one concept assertion
speaking about one individual



Published
Information
(an \mathcal{EL} Concept C)



Attacker's
Knowledge
(an $\mathcal{EL} / \mathcal{FL}_0 / \mathcal{FLE}$
Concept E)



Privacy Policy
(a finite set of
 \mathcal{EL} concepts)
 $\{D_1, \dots, D_p\}$

Formalizing Sensitive Information in \mathcal{EL} Instance Stores

- Given an \mathcal{EL} concept C (published information) and an \mathcal{EL} policy \mathcal{P}
- Given a quantifier symbol $Q \in \{\exists, \forall, \forall\exists\}$ and a DL
 $\mathcal{L}_{\exists} = \mathcal{EL}, \mathcal{L}_{\forall} = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FLE}$

- Given an \mathcal{EL} concept C (published information) and an \mathcal{EL} policy \mathcal{P}
- Given a quantifier symbol $Q \in \{\exists, \forall, \forall\exists\}$ and a DL
 $\mathcal{L}_\exists = \mathcal{EL}, \mathcal{L}_\forall = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FL}\mathcal{E}$

Compliance, Safety, Optimality

1. the \mathcal{L}_Q concept C' is **compliant with** \mathcal{P} if $C' \not\sqsubseteq D_i$ for all $i = 1, \dots, p$,

- Given an \mathcal{EL} concept C (published information) and an \mathcal{EL} policy \mathcal{P}
- Given a quantifier symbol $Q \in \{\exists, \forall, \forall\exists\}$ and a DL
 $\mathcal{L}_\exists = \mathcal{EL}, \mathcal{L}_\forall = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FLE}$

Compliance, Safety, Optimality

1. the \mathcal{L}_Q concept C' is **compliant with** \mathcal{P} if $C' \not\sqsubseteq D_i$ for all $i = 1, \dots, p$,
2. the \mathcal{EL} concept C' is
 - **Q -safe for** \mathcal{P} if for all \mathcal{L}_Q concepts E (attackers' knowledge) that are compliant with \mathcal{P} , $C' \sqcap E$ is also compliant with \mathcal{P} ,

- Given an \mathcal{EL} concept C (published information) and an \mathcal{EL} policy \mathcal{P}
- Given a quantifier symbol $Q \in \{\exists, \forall, \forall\exists\}$ and a DL
 $\mathcal{L}_\exists = \mathcal{EL}, \mathcal{L}_\forall = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FLE}$

Compliance, Safety, Optimality

1. the \mathcal{L}_Q concept C' is **compliant with** \mathcal{P} if $C' \not\sqsubseteq D_i$ for all $i = 1, \dots, p$,
2. the \mathcal{EL} concept C' is
 - **Q -safe for** \mathcal{P} if for all \mathcal{L}_Q concepts E (attackers' knowledge) that are compliant with \mathcal{P} , $C' \sqcap E$ is also compliant with \mathcal{P} ,
 - a **Q -safe generalization** of C for \mathcal{P} if $C \sqsubseteq C'$ and C' is Q -safe for \mathcal{P} ,

- Given an \mathcal{EL} concept C (published information) and an \mathcal{EL} policy \mathcal{P}
- Given a quantifier symbol $Q \in \{\exists, \forall, \forall\exists\}$ and a DL
 $\mathcal{L}_\exists = \mathcal{EL}, \mathcal{L}_\forall = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FLE}$

Compliance, Safety, Optimality

1. the \mathcal{L}_Q concept C' is **compliant with** \mathcal{P} if $C' \not\sqsubseteq D_i$ for all $i = 1, \dots, p$,
2. the \mathcal{EL} concept C' is
 - **Q -safe for** \mathcal{P} if for all \mathcal{L}_Q concepts E (attackers' knowledge) that are compliant with \mathcal{P} , $C' \sqcap E$ is also compliant with \mathcal{P} ,
 - a **Q -safe generalization** of C for \mathcal{P} if $C \sqsubseteq C'$ and C' is Q -safe for \mathcal{P} ,
 - an **optimal Q -safe generalization** of C for \mathcal{P} if
 - C' is a Q -safe generalization of C for \mathcal{P} and
 - there is no Q -safe generalization C'' of C for \mathcal{P} s.t. $C'' \sqsubset C'$.

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q -safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Modification



$C_1 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note $C \sqsubseteq C_1$ and C_1 **complies with** D

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

\exists -Attacker is Coming!



$C_1 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$



He knows $Patient(LINDA)$

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Linked and Revealed!



$C_1 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$
 \sqcap **Patient**

Note $D(LINDA)$ is **revealed**

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Modification



$C_2 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.\top)$
 $\sqcap \exists seenBy.(Male \sqcap worksIn.Oncology)$

C_2 is the (unique) **optimal \exists -safe generalization** for D

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

\forall -Attacker is Coming!



$C_2 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.T)$
 $\sqcap \exists seenBy.(Male \sqcap worksIn.Oncology)$



He knows $(Patient \sqcap \forall seenBy.\forall worksIn.Oncology)(LINDA)$

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$

$\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Linked and Revealed!



$C_2 = Female \sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.T)$

$\sqcap \exists seenBy.(Male \sqcap worksIn.Oncology)$

$\sqcap Patient \sqcap \forall seenBy.\forall worksIn.Oncology$

$D(LINDA)$ is revealed again

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q -safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Modification



$C_3 = Female \sqcap Patient \sqcap \exists seenBy.(Doctor \sqcap Male)$

Note C_3 is an **optimal \forall -safe generalization** for D

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

$\forall\exists$ -**Attacker is Coming!**



$C_3 = Female \sqcap Patient \sqcap \exists seenBy.(Doctor \sqcap Male)$



He knows $(\forall seenBy.\exists worksIn.Oncology)(LINDA)$

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q-safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Linked and Revealed!



$C_3 = Female \sqcap Patient \sqcap \exists seenBy.(Doctor \sqcap Male)$
 $\sqcap \forall seenBy.\exists worksIn.Oncology$

$D(LINDA)$ is **revealed again**

An Illustration on Privacy Attacks in \mathcal{EL} Instance Stores

Privacy Policy $\mathcal{P} = \{D\}$ about *LINDA*



$D = Patient \sqcap \exists seenBy.(Doctor \sqcap \exists worksIn.Oncology)$

Original **Published Information** C about *LINDA*



$C = Patient \sqcap Female$
 $\sqcap \exists seenBy.(Doctor \sqcap Male \sqcap \exists worksIn.Oncology)$

Note C is not **compliant with** and **Q -safe** for D for $Q \in \{\exists, \forall, \forall\exists\}$

Modification



$C_4 = Female$

Note C_4 is the **optimal $\forall\exists$ -safe generalization** for D !

Our Decision and Computational Problems

Given $Q \in \{\forall, \forall\exists\}$, a published information (\mathcal{EL} concept) C , an \mathcal{EL} policy \mathcal{P} .

Decision Problems

- **Q-Safety:**
Is an \mathcal{EL} concept C_1 Q-safe for a policy \mathcal{P} ?
- **Q-Optimality:**
Is an \mathcal{EL} concept C_1 an optimal Q-safe generalization of C for \mathcal{P} ?

Computational Problem

Find an \mathcal{EL} concept C_1 s.t C_1 is an optimal Q-safe generalization of C for \mathcal{P} !

compliance, \exists -**safety** and \exists -**optimality** have been investigated by (Baader, Kriegel, Nuradiansyah in JELIA 2019)

Complexity Results

Decision Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Q -safety	PTime*	PTime	PTime
Q -optimality	coNP* and Dual-hard*	coNP and Dual-hard	PTime

Table: Complexity results of decision problems on PPOP for \mathcal{EL} instance stores

Computational Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Optimal Q -safe Generalization(s)	ExpTime*	ExpTime	PTime

Table: Complexity of computing one/all optimal Q -safe generalizations for \mathcal{P}

* investigated by (Baader, Kriegel, and Nuradiansyah in JELIA 2019)

Decision Problems	$Q = \forall$	$Q = \forall$	$Q = \forall\exists$
Q -safety	PTime*	PTime	PTime
Q -optimality	coNP* and Dual-hard*	coNP and Dual-hard	PTime

Computational Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Optimal Q -safe Generalization(s)	ExpTime*	ExpTime	PTime

Reasons:

- Given an \mathcal{EL} concept D , $\text{con}(D)$ is the **set of all atoms** (A or $\exists r.D'$) in the top-level conjunction of D .
- Computing all **minimal hitting sets** of $\text{con}(D_1), \dots, \text{con}(D_p)$, where $\mathcal{P} = \{D_1, \dots, D_p\}$.
- The computation is performed recursively on the **role depth** of the published information C

Complexity Results

Decision Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Q -safety	PTime*	PTime	PTime
Q -optimality	coNP* and Dual-hard*	coNP and Dual-hard	PTime

Computational Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Optimal Q -safe Generalization(s)	ExpTime*	ExpTime	PTime

Reasons:

- Check if C_1 is an \forall -safe generalization of C for \mathcal{P}
- Check if there is C_2 s.t. $C \sqsubseteq C_2 \sqsubset C_1$, where C_2 is a not \forall -safe generalization of C for \mathcal{P}
- There is an NP algorithm to **guess** such concept C_2 (Baader, Kriegel, Nuradiansyah in JELIA 2019)

Decision Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Q-safety	PTime*	PTime	PTime
Q-optimality	coNP* and Dual-hard*	coNP and Dual-hard	PTime

Computational Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Optimal Q-safe Generalization(s)	ExpTime*	ExpTime	PTime

Reasons:

- \forall -Optimality is coNP-hard? Don't know yet
- There is a polynomial reduction of Dual problem to \forall -optimality

Given two **families of inclusion-comparable sets** \mathcal{G} and \mathcal{H} , Dual asks whether \mathcal{H} consists exactly of the minimal hitting sets of \mathcal{G} .

Decision Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Q-safety	PTime*	PTime	PTime
Q-optimality	coNP* and Dual-hard*	coNP and Dual-hard	PTime

Computational Problems	$Q = \exists$	$Q = \forall$	$Q = \forall\exists$
Optimal Q-safe Generalization(s)	ExpTime*	ExpTime	PTime

Reasons:

$\forall\exists$ -Safety and $\forall\exists$ -Optimality

C is $\forall\exists$ -safe for \mathcal{P} iff

1. $A \notin \text{con}(C)$ for all concept names $A \in \text{con}(D_1) \cup \dots \cup \text{con}(D_p)$, and
2. for all existential restrictions $\exists r.D' \in \text{con}(D_1) \cup \dots \cup \text{con}(D_p)$, there is no concept of the form $\exists r.E \in \text{con}(C)$

Conclusions:

- Investigate **PPOP for \mathcal{EL} Instance Stores**
- Considering **attacker's knowledge** to be given by an \mathcal{FL}_0 or \mathcal{FLE} concept
- Deciding **Q -safety** and **Q -optimality**, where $Q \in \{\forall, \exists\}$.
- Computing **optimal Q -safe generalizations** of \mathcal{EL} concepts for \mathcal{P}

Note: *the stronger the attacker's knowledge, the more radical we need to change the concept to make it safe*

Conclusions:

- Investigate **PPOP for \mathcal{EL} Instance Stores**
- Considering **attacker's knowledge** to be given by an \mathcal{FL}_0 or $\mathcal{FL}\mathcal{E}$ concept
- Deciding **Q -safety** and **Q -optimality**, where $Q \in \{\forall, \forall\exists\}$.
- Computing **optimal Q -safe generalizations** of \mathcal{EL} concepts for \mathcal{P}

Note: *the stronger the attacker's knowledge, the more radical we need to change the concept to make it safe*

Future Work:

- PPOP in \mathcal{EL} ABoxes, including role assertions (Ongoing!)
- PPOP in \mathcal{EL} Instance Stores w.r.t. (General) TBoxes
- Playing with more different or expressive DLs

Thank You

ROSI