technische universität dortmund

Information Control by Policy-Based Relational Weakening Templates

Joachim Biskup **Marcel Preuß**

Information Systems and Security (ISSI)

Technische Universität Dortmund, Germany

September 30, 2016

technische universität dortmund

Context of this Work

Inference-Proof Data Publishing

Nowadays: Data publishing is ubiquitous

- Governments and companies provide data
- People share data about their private lives

But: Original data often contains sensitive (personal) information

- Set up a confidentiality policy
- Release "secure views" instead of original data
  - Do not reveal any confidential information
  - Consider adversary's abilities to infer information

Framework and Goal (Inference-Proofness)

Framework: Relational model relying on first-order logic

- Complete original instance r (definite knowledge: $+/-$)
- Confidentiality policy *ppol* of prohibitions $(\exists \boldsymbol{X})\, R(\boldsymbol{X}, \boldsymbol{c})$ s.t. each variable X occurs only once
- Adversary is aware of policy and protection mechanism

Goal: Enforce policy **efficiently** by weakened view on r s.t.

- Weakened view $weak(r, ppol)$ contains only true knowledge
- Inference-proofness from adversary's point of view:
  For each $\Psi \in ppol$ there is a "secure" alternative instance r^{Ψ}
  - r^{Ψ} does **not satisfy** Ψ
  - r^{Ψ} is **indistinguishable** from original instance r
    $\rightarrow weak(r^{\Psi}, ppol) = weak(r, ppol)$

Confidentiality by Weakening

technische universität dortmund

Construction of Weakened Views

Stage 1: Disjoint disjunction templates *(independent of r)*

- Partition the policy *ppol* into disjoint clusters $C_1, \ldots, C_q$ (inducing disjunction templates) of a certain minimum size
- If necessary: Construct additional prohibitions

Stage 2: Weakened view $weak(r, ppol)$ *(dependent on r)*

- Keep each tuple of r not satisfying any $\Psi \in C_i$
- Introduce each disjunction $\bigvee_{\Psi \in C_i} \Psi$ satisfied by r
- Knowledge not satisfying kept tuples or disjuncts is negative

$\rightarrow$ Three classes of knowledge: $+$, $\vee$, $-$

Confidentiality by Weakening: Example (1)

Policy: $ppol = \{\ \Psi_1 = R(a,b,c),\ \ \Psi_2 = R(a,b,d)\ \}$

Complete original instance r:

+	−
(a,b,c)	(a,a,a)
(a,c,c)	(a,a,b)
(b,a,c)	$\vdots$
	(a,b,d)
	$\vdots$

$\Longrightarrow$

$$R(a,b,c),\ R(a,c,c),\ R(b,a,c)$$

$$(\forall X)(\forall Y)(\forall Z)\ [\ (X \equiv a \wedge Y \equiv b \wedge Z \equiv c) \vee (X \equiv a \wedge Y \equiv c \wedge Z \equiv c) \vee (X \equiv b \wedge Y \equiv a \wedge Z \equiv c) \vee \neg R(X,Y,Z)\ ]$$

Obviously: r satisfies Ψ_1 ($\rightarrow$ to be weakened)

Confidentiality by Weakening: Example (2)

Disjunction template: $\Psi_1 \vee \Psi_2 \; = \; R(a,b,c) \; \vee \; R(a,b,d)$

Weakened view $weak(r, ppol)$:

+	−
~~(a,b,c)~~	(a,a,a)
(a,c,c)	(a,a,b)
(b,a,c)	$\vdots$
	~~(a,b,d)~~
	$\vdots$

$\Longrightarrow$

$R(a,c,c),\; R(b,a,c)$

$R(a,b,c) \vee R(a,b,d)$

$$(\forall X)(\forall Y)(\forall Z)\,[\; (X \equiv a \wedge Y \equiv b \wedge Z \equiv c) \vee (X \equiv a \wedge Y \equiv b \wedge Z \equiv d) \vee (X \equiv a \wedge Y \equiv c \wedge Z \equiv c) \vee (X \equiv b \wedge Y \equiv a \wedge Z \equiv c) \vee \neg R(X,Y,Z) \;]$$

Disjunctive knowledge:
$R(a,b,c) \vee R(a,b,d)$

Achievement: $weak(r, ppol)$ does **neither** imply Ψ_1 **nor** Ψ_2

Inference-Proofness by Isolation

Structure of weakened views:

$+$ $R(\boldsymbol{c}_1),\ R(\boldsymbol{c}_2),\ \ldots,\ R(\boldsymbol{c}_p)$ (definite knowl.)

↓ $R(\boldsymbol{c}_i) \not\models_{DB} \Psi_{j,\ell}$

$\vee$ $\Psi_{1,1} \vee \ldots \vee \Psi_{1,k_1} \quad \ldots \quad \Psi_{m,1} \vee \ldots \vee \Psi_{m,k_m}$ ⟲ $\Psi_{i,j} \not\models_{DB} \Psi_{\bar{i},\bar{j}}$

↑ $\neg R(\boldsymbol{d}_i) \not\models_{DB} \neg\Psi_{j,\ell}$

$-$ $\neg R(\boldsymbol{d}_1),\ \neg R(\boldsymbol{d}_2),\ \neg R(\boldsymbol{d}_3),\ \ldots$ (definite knowl.)

Hence: For each $\Psi \in \Psi_{i,1} \vee \ldots \vee \Psi_{i,k_i}$ alternative instance r^{Ψ} with

- $r^{\Psi} \not\models_M \Psi$ ✓ (but: $r^{\Psi} \models_M \Psi_{i,1} \vee \ldots \vee \Psi_{i,k_i}$)
- $r^{\Psi} \models_M +, \vee, -$ $\rightsquigarrow$ indistinguishability by construction of weakened views ✓

About the Clustering of Policy Elements

Desired properties for disjoint disjunction templates

- Credibility of all disjuncts $\rightsquigarrow$ confidentiality
- Semantically meaningful $\rightsquigarrow$ availability
- Certain length $\rightsquigarrow$ level of confidentiality/availability

Desired properties for disjoint clustering of policy elements

- Consider (high-level) specification of admissible clusters
  $\rightarrow$ Depends on application scenario
- Each cluster must have a certain (minimum) size k^*
- Minimize number of additional prohibitions

Clustering problem is NP-hard for $k^* \geq 3$ (Reduction of X3C)

Inference-Proofness under A Priori Knowledge

Introducing A Priori Knowledge

Usually: Adversary also has some a priori knowledge *prior*

Challenge for inference-proofness: "secure" alternative instance r^{Ψ}

- r^{Ψ} does **not satisfy** Ψ } (already discussed)
- r^{Ψ} is **indistinguishable** from original r } (already discussed)
- r^{Ψ} **satisfies** *prior*

Assumed *prior*: "Single Premise TGDs" of the form

$$\Gamma \;:=\; (\forall \boldsymbol{X})\,[\,R(\boldsymbol{X}, \boldsymbol{c}_1) \Rightarrow (\exists \boldsymbol{Y})\,R(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{c}_2)\,] \quad \text{s.t.}$$

- each X occurs only once in $prem(\Gamma)$ and
- each X, Y occurs only once in $concl(\Gamma)$

Confidentiality Compromising Dependencies

Semantics of Single Premise TGDs: (also via transitive chains)

- Existent DB-Tuple ⇒ Existence of other DB-Tuple
- Non-Existent DB-Tuple ⇒ Non-Existence of other DB-Tuple

Broken isolation in weakened views:

$+$ $R(\boldsymbol{c}_1),\ R(\boldsymbol{c}_2),\ \ldots,\ R(\boldsymbol{c}_p)$ (definite knowl.)

Dependencies

$\vee$ $\Psi_{1,1} \vee \ldots \vee \Psi_{1,k_1} \quad \ldots \quad \Psi_{m,1} \vee \ldots \vee \Psi_{m,k_m}$ Dependencies

Dependencies

$-$ $\neg R(\boldsymbol{d}_1),\ \neg R(\boldsymbol{d}_2),\ \neg R(\boldsymbol{d}_3),\ \ldots$ (definite knowl.)

Re-Establishing Sufficient Isolation (1)

Handling of dependency Γ interfering with policy elements

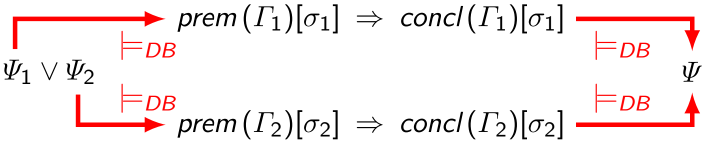
- Add policy elements protecting *prem*(Γ) and *concl*(Γ)
  $\rightarrow$ Do not reveal satisfaction-status of premise or conclusion
- Attention: New policy elements $\rightsquigarrow$ further interferences

Problem: Distortions by disjunctions do not always guarantee
the possibility of non-satisfaction of conclusions

Only escape: Resort to distortion by complete refusal ☹

Re-Establishing Sufficient Isolation (2)

Inference-channel within disjunctive knowledge:



How to eliminate this kind of inference-channel?

- Partitioning of *prior* s.t. Γ_1 and Γ_2 in same partition, if
  - their conclusions imply the same Ψ (under some σ_1, σ_2) or
  - they can possibly form a transitive chain
- Do not construct disjunction, if
  each disjunct implies a premise of the same partition

technische universität dortmund

Experimental Evaluation and Conclusion

Experimental Evaluation for $k^* = 2$

Lessons learned from 5 experiment setups

- Algorithm efficiently handles input instances of realistic size
  - 1 000 000 database tuples
  - 100 000 policy elements
  - 2500 dependencies

  $\Big\}$ $\rightsquigarrow$ 1 minute
- Size and structure of *ppol* and *prior* crucial for runtime
- Parallelization: Doubling threads nearly halves runtime

technische universität dortmund

Conclusion & Future Work

Main contributions:

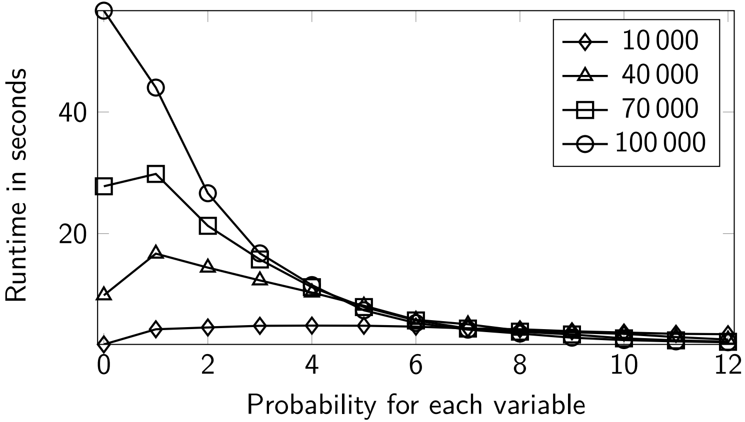
- Confidentiality by cooperative weakening without lies
- Even if adversary employs Single Premise TGDs
- Efficient computation for disjunctions of length $k^* = 2$

Possible future work:

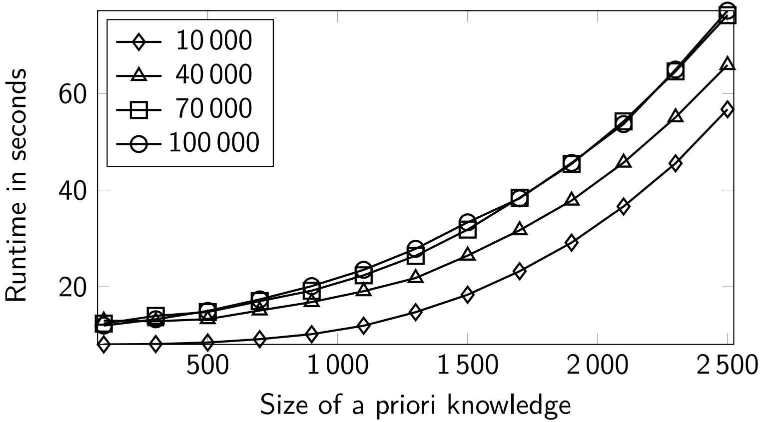
- Clustering algorithm for $k^* \geq 3$ $\quad$ ($\rightarrow$ Reasonable heuristic)
- More expressive classes of a priori knowledge
- Model k-anonymity/ℓ-diversity within weakening approach

Experimental Results

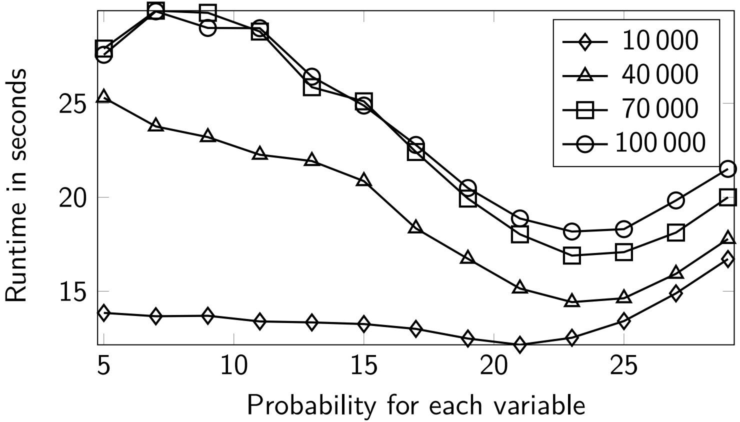
Experiment: Varying Number of Variables in Policies



Experiment: Varying Number of Dependencies



Experiment: Varying Number of Variables in *prior*



technische universität dortmund

Experiment: Varying Number of Threads