technische universität dortmund

On the Inference-Proofness of Database Fragmentation Satisfying Confidentiality Constraints

Marcel Preuß

Chair VI (ISSI), Computer Science

Technische Universität Dortmund

March 31, 2011

Table of Contents

Confidentiality by Fragmentation

Information as a Ressource

Today: Information is an important ressource
$\rightarrow$ Confidentiality of information is important

Economy-Driven society: Cost-efficiency of importance
$\rightarrow$ Outsourcing: "Database as a service"-Paradigm

Goal conflict: Confidentiality $\leftrightsquigarrow$ Outsourcing

Approaches to Achieving Confidentiality

Confidentiality by encryption on user-side?
→ Efficient handling of queries on server-side difficult

Often: Only associations between pieces of information sensitive

Example: Situation in a hospital

- List of illnesses cured ⇝ Not sensitive
- List of patients ⇝ Not really sensitive
- Association: Patient and his illness → Very sensitive

Confidentiality by Fragmentation: Example (1)

Patient	SSN	Name	DoB	ZIP	Illness	Doctor
	12345	Hellmann	03.01.1981	94142	Hypertension	White
	98765	Dooley	07.10.1953	94141	Obesity	Warren
	24689	McKinley	12.02.1952	94142	Hypertension	White
	13579	Ripley	03.01.1981	94139	Obesity	Warren

Figure: Instance *patient* over schema *Patient*

Noticeable

- Attribute `SSN` is a primary key
- Sensitive associations are contained

Confidentiality by Fragmentation: Example (2)

F_1	Name
	Hellmann
	Dooley
	McKinley
	Ripley

F_2	DoB	ZIP
	03.01.1981	94142
	07.10.1953	94141
	12.02.1952	94142
	03.01.1981	94139

F_3	Illness	Doctor
	Hypertension	White
	Obesity	Warren

Figure: Possible fragment instances of *patient*

Noticeable

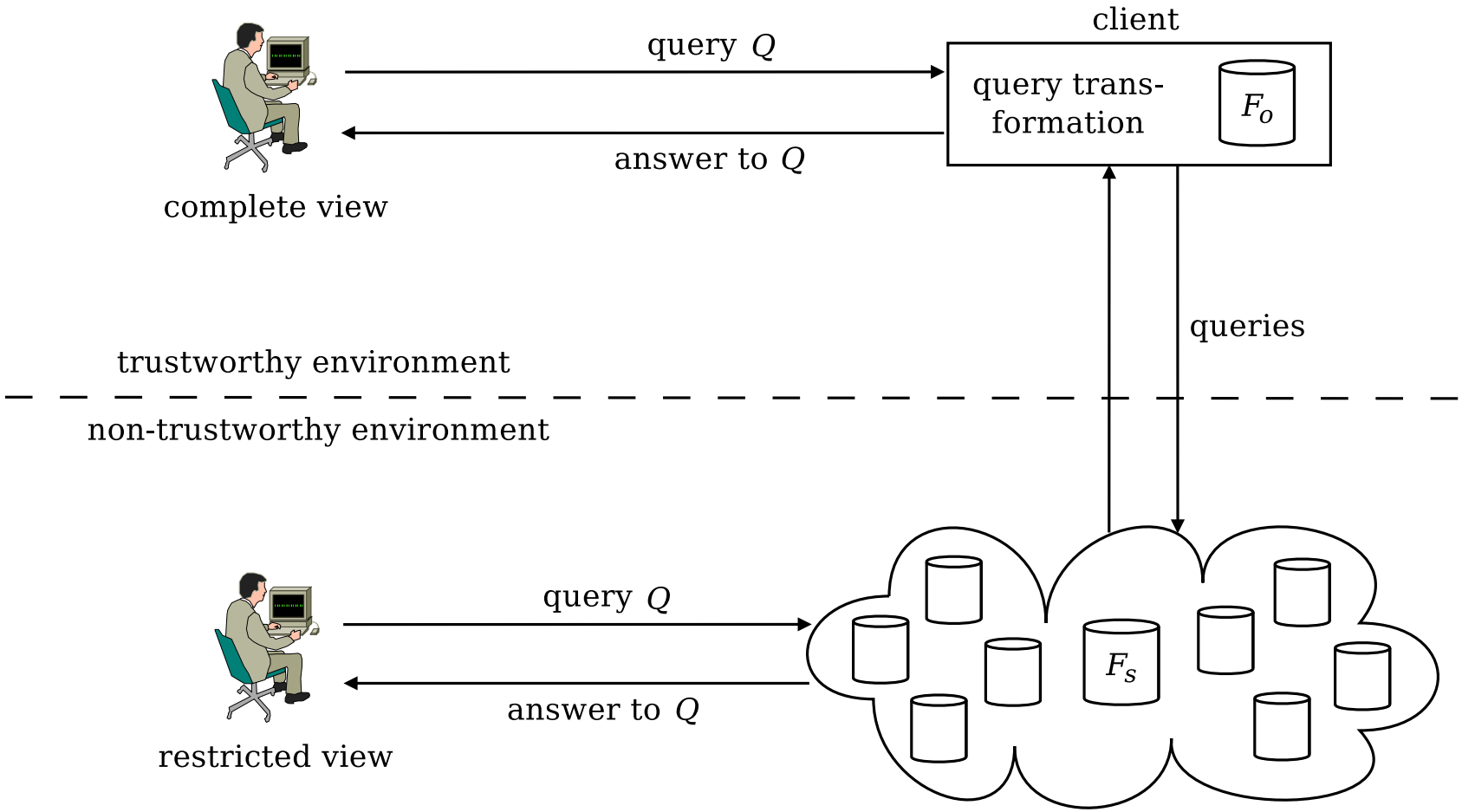
- Primary key `SSN` not in any fragment
- Sensitive associations broken

Towards an Approach to Fragmentation

Assumptions: Underlying client-server framework

- Server is honest, but curious
- Client is completely trustworthy
- Client has (limited) local storage
- Local storage more expensive than external storage
  $\rightarrow$ Target: Use external storage for as much data as possible

Working with Fragmented Databases



Fragmentation Compliant with Assumptions

Fragmentation of instance r over schema $\langle R|A_R|SC_R\rangle$

- Fragmentation on schema level
  - Set of Fragments $\mathcal{F} = \{\langle F_o|A_{F_o}|SC_{F_o}\rangle, \langle F_s|A_{F_s}|SC_{F_s}\rangle\}$
  - $\langle F_i|A_{F_i}|SC_{F_i}\rangle$ is a relational schema with $A_{F_i} \subseteq A_R$
  - Each attribute of A_R is contained in exactly one fragment
- Fragmentation on instance level
  - Fragment instances f_o and f_s: Projections of r on A_{F_o} and A_{F_s}
  - Local storage of instance f_o $(\rightarrow$ Owner$)$
  - External storage of instance f_s $(\rightarrow$ Server$)$

Example of a Possible Fragmentation

F_o	SSN	Name	DoB
	12345	Hellmann	03.01.1981
	98765	Dooley	07.10.1953
	24689	McKinley	12.02.1952
	13579	Ripley	03.01.1981

F_s	ZIP	Illness	Doctor
	94142	Hypertension	White
	94141	Obesity	Warren
	94139	Obesity	Warren

Figure: Possible fragmentation of *patient*

Considering Reconstructability

Problem: Reconstructability of r not guaranteed

Idea: Usage of Tuple-Identifiers (T-IDs)

- Add attribute $\texttt{tid} \notin A_R$ to both A_{F_o} and A_{F_s} as a primary key
- In both f_o and f_s:
  - Tuples belonging together have a unique T-ID in common
  - Consequence: Duplicates regarding $A_{F_i} \setminus \{\texttt{tid}\}$ are kept

Example of a Possible Fragmentation with T-IDs

F_o	tid	SSN	Name	DoB
	1	12345	Hellmann	03.01.1981
	2	98765	Dooley	07.10.1953
	3	24689	McKinley	12.02.1952
	4	13579	Ripley	03.01.1981

F_s	tid	ZIP	Illness	Doctor
	1	94142	Hypertension	White
	2	94141	Obesity	Warren
	3	94142	Hypertension	White
	4	94139	Obesity	Warren

Figure: Possible fragmentation of *patient* with T-IDs

Formal Declaration of Confidentiality Requirements

How to declare confidentiality requirements formally?

Confidentiality Constraint c over $\langle R|A_R|SC_R\rangle$ is a subset $c \subseteq A_R$

Correctness of $\mathcal{F} = \{\langle F_o|A_{F_o}|SC_{F_o}\rangle, \langle F_s|A_{F_s}|SC_{F_s}\rangle\}$:

- Let $\mathcal{C}$ be a set of Confidentiality Constraints
- $\mathcal{F}$ is correct w.r.t. $\mathcal{C} \quad \Leftrightarrow \quad c \nsubseteq A_{F_s}$ holds for all $c \in \mathcal{C}$

Example: Set of Confidentiality Constraints

$$c_0 = \{\texttt{SSN}\}$$
$$c_1 = \{\texttt{Name}, \texttt{DoB}\}$$
$$c_2 = \{\texttt{Name}, \texttt{ZIP}\}$$
$$c_3 = \{\texttt{Name}, \texttt{Illness}\}$$
$$c_4 = \{\texttt{Name}, \texttt{Doctor}\}$$
$$c_5 = \{\texttt{DoB}, \texttt{ZIP}, \texttt{Illness}\}$$
$$c_6 = \{\texttt{DoB}, \texttt{ZIP}, \texttt{Doctor}\}$$

Figure: Set $\mathcal{C}$ of Confidentiality Constraints over *Patient*

Example: Correct Fragmentation

F_o	tid	SSN	Name	DoB
	1	12345	Hellmann	03.01.1981
	2	98765	Dooley	07.10.1953
	3	24689	McKinley	12.02.1952
	4	13579	Ripley	03.01.1981

F_s	tid	ZIP	Illness	Doctor
	1	94142	Hypertension	White
	2	94141	Obesity	Warren
	3	94142	Hypertension	White
	4	94139	Obesity	Warren

Figure: Fragmentation of *patient*, correct w.r.t. $\mathcal{C}$

Inference-Proofness of Fragmentation

Approach to Show Inference-Proofness

How to succeed in analysing inference-proofness?

- CQE is known to be inference-proof
- Modelling of fragmentation within the CQE-Framework
  - Choice of an appropriate logic
  - Modelling of f_s, r and their relationship
  - Modelling of confidentiality constraints
- Formal proof within logic-oriented framework
  - Assumptions about an attacker's reasoning abilites
  - Assumptions about an attacker's a priori knowlegde

Choice of an Appropriate Logic: Syntax

Syntax of the logic ($\rightarrow$ Language $\mathscr{L}$)

- 1st-order logic with equality
  - Predicate symbol R with arity n
  - Predicate symbol F_s with arity k
  - Distinguished binary predicate symbol $=$
  - Fixed infinite domain Dom
    $\rightarrow$ Constant symbols declared for the relational schema
  - Infinite set of variables $Var := \{X_1, X_2, \ldots\}$
- Only constants or variables as terms of atomic formulas
- Only closed formulas $\rightarrow$ All variables are quantified ($\forall, \exists$)

Choice of an Appropriate Logic: Semantics

An interpretation $\mathcal{I}$ for $\mathscr{L}$ is a DB-Interpretation $\Leftrightarrow$

- Universe $\mathcal{U}$ = Domain Dom
- $\mathcal{I}(v) = v$ holds for all $v \in Dom$
- R is interpreted by a finite set $\mathcal{I}(R) \subset \mathcal{U}^n$
- F_s is interpreted by a finite set $\mathcal{I}(F_s) \subset \mathcal{U}^k$
- For predicate symbol $=$ holds: $\mathcal{I}(=) = \{ (v, v) \mid v \in \mathcal{U} \}$

Implication Based on DB-Interpretation

Notion of Satisfaction

- Consider a DB-Interpretation $\mathcal{I}$
- Set of formulas $\mathcal{S} \subset \mathscr{L}$
- $\mathcal{I}$ satisfies $\mathcal{S}$ is written as $\mathcal{I} \models_M \mathcal{S}$

Semantics of satisfaction: The same as in usual first-order logic

$\mathcal{S} \subset \mathscr{L}$ implies $\Phi \in \mathscr{L}$ (written $\mathcal{S} \models_{DB} \Phi$) iff
For each DB-Interpretation $\mathcal{I}$ with $\mathcal{I} \models_M \mathcal{S}$ also $\mathcal{I} \models_M \Phi$ holds

Convention from now on

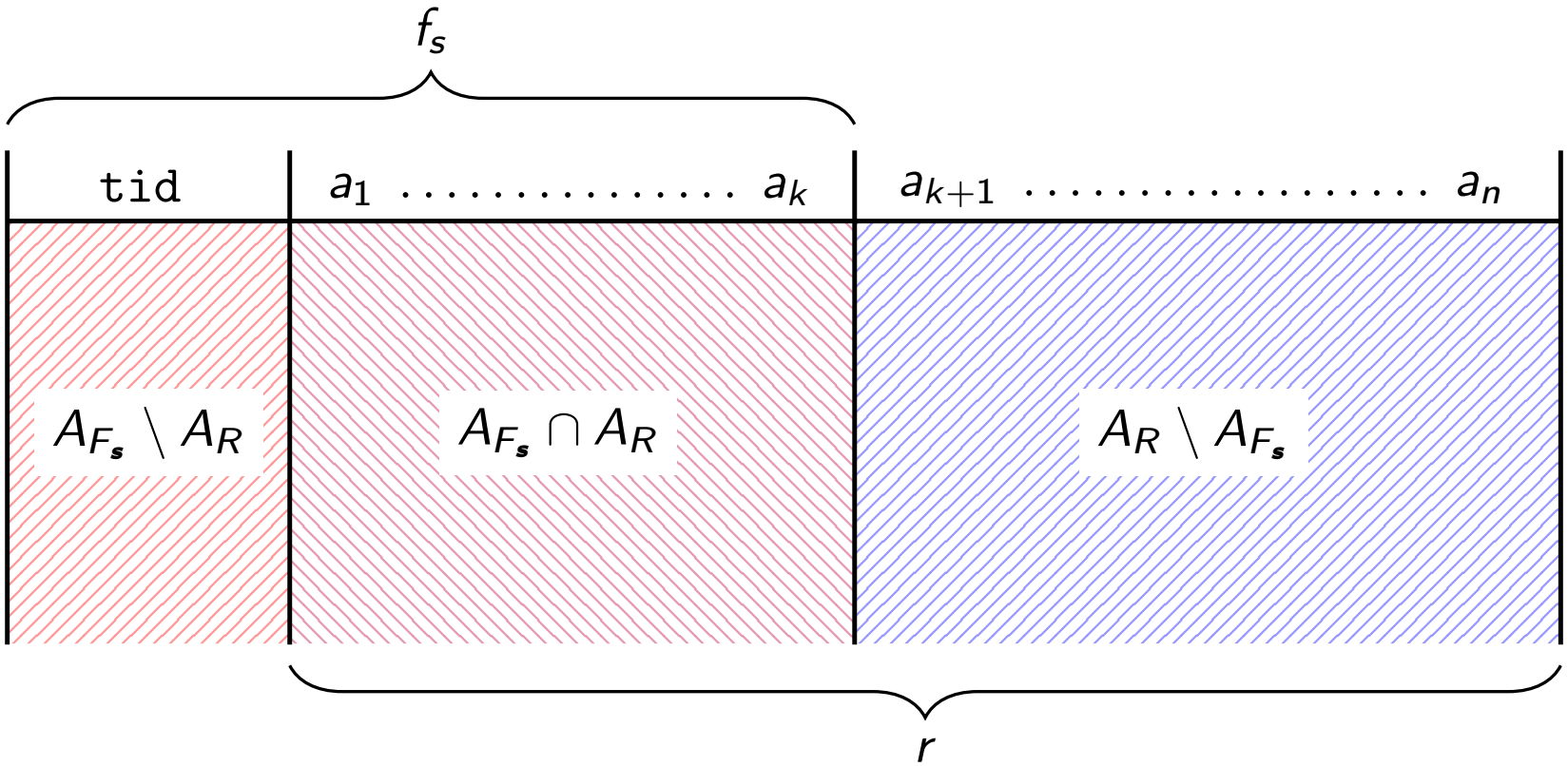


Figure: Convention (w.l.o.g.): Rearrangement of columns of r

Modelling the Positive Knowledge of f_s

An attacker knows about the visible fragment

- Outsourced fragment instance f_s
- Fragment $\langle F_s | A_{F_s} | SC_{F_s} \rangle$ with $A_{F_s} = \{a_{\mathtt{tid}}, a_1, \ldots, a_k\}$

Explicit positive knowlegde of f_s from an attacker's point of view

- $db_{f_s}^{+} := \{F_s(\nu[a_{\mathtt{tid}}], \nu[a_1], \ldots, \nu[a_k]) \mid \nu \in f_s\}$
- Functional dependency $a_{\mathtt{tid}} \rightarrow \{a_1, \ldots, a_k\} \in SC_{F_s}$

Example of Modelling the Positive Knowledge of f_s

F_s	**tid**	**ZIP**	**Illness**	**Doctor**
	1	94142	Hypertension	White
	2	94141	Obesity	Warren
	3	94142	Hypertension	White
	4	94139	Obesity	Warren

$$db^+_{f_s} = \{ \begin{array}{l} F_s(1, 94142, \text{Hypertension}, \text{White}), \\ F_s(2, 94141, \text{Obesity}, \text{Warren}), \\ F_s(3, 94142, \text{Hypertension}, \text{White}), \\ F_s(4, 94139, \text{Obesity}, \text{Warren}) \} \end{array}$$

Negative Knowledge Resulting from Completeness

Problem: An attacker knows even more about f_s

- Instances r and f_s are supposed to be complete
- Every constant combination not in f_s is invalid in f_s by CWA
  $\rightarrow$ Knowledge of the kind $\neg F_s(v_{\texttt{tid}}, v_1, \ldots, v_n)$
- Problem: Infinite Domain $\rightarrow$ Not explicitly enumerable
- Bright idea: Use Completeness-Sentence to model CWA

Modelling the Negative Knowledge of f_s

CWA in terms of the running example:

$$
\begin{array}{l}
(\forall X_t)(\forall X_Z)(\forall X_I)(\forall X_D)\ [\\
(X_t = 1 \ \wedge \ X_Z = 94142 \ \wedge \ X_I = \text{Hypert.} \ \wedge \ X_D = \text{White}) \ \vee \\
(X_t = 2 \ \wedge \ X_Z = 94141 \ \wedge \ X_I = \text{Obesity} \ \wedge \ X_D = \text{Warren}) \ \vee \\
(X_t = 3 \ \wedge \ X_Z = 94142 \ \wedge \ X_I = \text{Hypert.} \ \wedge \ X_D = \text{White}) \ \vee \\
(X_t = 4 \ \wedge \ X_Z = 94139 \ \wedge \ X_I = \text{Obesity} \ \wedge \ X_D = \text{Warren}) \ \vee \\
\neg F_s(X_t, X_Z, X_I, X_D) \ ]
\end{array}
$$

CWA as a Completeness Sentence in $db_{f_s}^{-}$:

$$
(\forall X_{\mathtt{tid}}) \ldots (\forall X_k) \left[\bigvee_{\nu \in f_s} \left(\bigwedge_{a_j \in A_{F_s}} (X_j = \nu[a_j]) \right) \vee \neg F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k) \right]
$$

Final Logic-Oriented view on f_s

Summing up: A logic-oriented view on f_s is modelled by

$$db_{f_s} := db^+_{f_s} \cup db^-_{f_s} \cup \{a_{\mathtt{tid}} \to \{a_1, \ldots, a_k\}\}$$

But: An attacker is interested in knowing the original instance r

The Knowledge Known About r (1)

Suppose: Attacker knows the process of fragmentation including

- Fragment instance f_s over $\langle F_s | A_{F_s} | SC_{F_s} \rangle$
- Schema $\langle R | A_R | SC_R \rangle$ over which r is built

Knowledge resulting from relationship between f_s and r

- For each $\nu \in f_s$: A tuple $\mu \in r$ with $\mu\lceil A_{F_s} = \nu\lceil A_R$ exists
- For each $\nu \notin f_s$: No tuple $\mu \in r$ with $\mu\lceil A_{F_s} = \nu\lceil A_R$

Knowledge expressed as a formula of db_r:

$$(\forall X_1) \ldots (\forall X_k) \left[(\exists X_{\mathtt{tid}})\, F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k) \Leftrightarrow (\exists X_{k+1}) \ldots (\exists X_n)\, R(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n) \right]$$

The Knowledge Known About r (2)

Knowledge resulting from unique T-IDs contained in f_s

- Duplicates of tuples regarding $A_{F_s} \cap A_R$ are kept
- But: Corresponding tuples in r cannot be equal

Knowledge expressed as a formula of db_r:

$$(\forall X_1) \ldots (\forall X_k) \left[(\exists X_{\mathtt{tid}}) (\exists X'_{\mathtt{tid}}) \left[F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k) \wedge F_s(X'_{\mathtt{tid}}, X_1, \ldots, X_k) \wedge (X_{\mathtt{tid}} \neq X'_{\mathtt{tid}}) \right] \Rightarrow \right.$$

$$\left. (\exists X_{k+1}) \ldots (\exists X_n) \left(\exists X'_{k+1}\right) \ldots (\exists X'_n) \left[R(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n) \wedge R(X_1, \ldots, X_k, X'_{k+1}, \ldots, X'_n) \wedge \bigvee_{j=k+1}^{n} (X_j \neq X'_j) \right] \right]$$

Confidentiality Constraints in the CQE-Framework

Design choice: Confidentiality constraints as potential secrets

- Supposition: Only those values or associations recorded in r are protected by confidentiality constraints
- About a potential secret $\Psi \in \mathscr{L}$ defined for a user:
  - If Ψ is true in instance: User must *not* get to know this
  - Otherwise: User may know that Ψ is false in instance
- Assume: An attacker is aware of $\mathcal{C}$

Bridging the Differences

From attribute-level to value-level

- Consider a confidentiality constraint $c_i = \{a_{i_1}, \ldots, a_{i_\ell}\}$
- Protect *all* constant combinations possible for $a_{i_1}, \ldots, a_{i_\ell}$
  $\rightarrow$ One potential secret per possible combination
- Otherwise: Attacker can read secrets directly from $pot\_sec(\mathcal{C})$
- But: Leads to an infinite number of formulas as $|Dom| = \infty$
- Idea: Upgrade $\mathscr{L} \rightarrow \mathscr{L}^f \supset \mathscr{L}$ containing free variables
- Use free variables $X_{i_1}, \ldots, X_{i_\ell}$ to represent $a_{i_1}, \ldots, a_{i_k}$

Modelling of Confidentiality Constraints

Consider a confidentiality constraint $c_i = \{a_{i_1}, \ldots, a_{i_\ell}\} \in \mathcal{C}$

- $\mathrm{Ind}^+_{c_i} = \{i_1, \ldots, i_\ell\}$
- $\mathrm{Ind}^-_{c_i} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_\ell\} = \{i_{\ell+1}, \ldots, i_n\}$

Construction of $pot\_sec(\mathcal{C})$:

- For all $c_i \in \mathcal{C}$: Add the potential secret

$$\Psi_i(\boldsymbol{X_i}) = (\exists X_{i_{\ell+1}}) \ldots (\exists X_{i_n})\, R(X_1, \ldots, X_n)$$

- Thereby, for $j \in \{1, \ldots, n\}$:
  - If $j \in \mathrm{Ind}^+_{c_i}$: X_j is a free variable
  - If $j \in \mathrm{Ind}^-_{c_i}$: X_j is a quantified variable
- $\boldsymbol{X_i} = (X_{i_1}, \ldots, X_{i_\ell})$ is the vector of free variables

Expansion of the Confidentiality Policy

Given: $\Psi_i(\boldsymbol{X_i})$ with $\boldsymbol{X_i} = (X_{i_1}, \ldots, X_{i_\ell})$

Problem: Semantics for $\mathscr{L}$ does not comprise free variables

Solution: Construction of Expansion $\mathrm{ex}(\Psi_i(\boldsymbol{X_i})) \subset \mathscr{L}$

- Consider each constant combination $\boldsymbol{v_i} = (v_{i_1}, \ldots, v_{i_\ell})$
- Construct each formula $\Psi_i(\boldsymbol{v_i}) \in \mathrm{ex}(\Psi_i(\boldsymbol{X_i}))$

Expansion of $pot\_sec(\mathcal{C})$:

$$\mathrm{ex}(pot\_sec(\mathcal{C})) := \bigcup_{\Psi(\boldsymbol{X}) \in pot\_sec(\mathcal{C})} \mathrm{ex}(\Psi(\boldsymbol{X}))$$

About A-Priori Knowledge

Known now

- Logic-oriented view on fragmentation
- Until now: An attacker's a priori knowledge is neglected

Prior work: A priori knowledge of crucial importance

- Fragmentation already known to be inference-proof, if
  - No a priori knowledge
  - A priori knowledge in terms of functional dependencies
- Not inference-proof under general a priori knowledge

Now: Inference-proofness under unirelational typed EGDs/TGDs

About Unirelational EGDs/TGDs

Considered: Semantic constraints SC_R of $\langle R|A_R|SC_R\rangle$

Nearly all semantic constraints can be characterized as

- Equality Generating Dependencies (EGDs) (e.g. FDs)
- Tuple Generating Dependencies (TGDs) (e.g. JDs, INDs)

Unirelational EGD/TGD: $(\forall \boldsymbol{X})\,[\,\alpha(\boldsymbol{X}) \Rightarrow (\exists \boldsymbol{Y})\,\beta(\boldsymbol{X}, \boldsymbol{Y})\,]$ with

- α is a conjunction of atoms $R(\ldots)$ over variables of $\boldsymbol{X}$
- β is a conjunction of atoms $R(\ldots)$ and $(\ldots = \ldots)$ over $\boldsymbol{X}$, $\boldsymbol{Y}$
- All variables of $\boldsymbol{X}$ appear in α
- All terms are variables ($\rightarrow$ No constants allowed!)

About Typed Constraints

Typed EGD/TGD: *Var* can be partitioned into n disjoint classes:

- For each atom $R(X_1, \ldots, X_n)$: X_i in class i
- For each atom $(X' = X'')$: X' and X'' belong to the same class

Examples of (un)typed EGDs/TGDs

- $(\forall \boldsymbol{X})\,[\,R(X_1, X_2, X_1, \ldots) \Rightarrow R(\ldots)\,]$
- $(\forall \boldsymbol{X})\,[\,R(X_1, X_2, X_3, \ldots) \Rightarrow (X_1 = X_2)\,]$
- $(\forall \boldsymbol{X})\,[\,R(X_1, X_2, X_3, \ldots) \wedge R(X_1, X_3, X_2' \ldots) \Rightarrow R(\ldots)\,]$
- $(\forall \boldsymbol{X})\,[\,R(X_1, X_2, X_3, \ldots) \wedge R(X_1, X_2', X_3', \ldots) \Rightarrow (X_3 = X_3')\,]$
- $(\forall \boldsymbol{X})\,[\,R(X_1, X_2, X_3, \ldots) \wedge R(X_1', X_2', X_3, \ldots) \Rightarrow R(X_1, X_2', \ldots)\,]$

Summary of Views on Fragmentation

Relational Level	**Logic-Oriented Level**
Instance r over $\langle R \vert A_R \vert SC_R \rangle$	Set of formulas db_r
Confidentiality Constraints $\mathcal{C}$	Confident. Policy $pot\_sec(\mathcal{C})$
Fragm. $\mathcal{F}$, correct w.r.t. $\mathcal{C}$	Implicitly in db_r
f_s over $\langle F_s \vert A_{F_s} \vert SC_{F_s} \rangle \in \mathcal{F}$	Set of formulas db_{f_s}
EGDs/TGDs in SC_R	A-Priori Knowledge $prior_{SC_R}$

Sketch of Proof

To be shown:
for all $\Psi(\boldsymbol{v}) \in \text{ex}(pot\_sec(\mathcal{C}))$: $db_{f_s} \cup db_r \cup prior_{SC_R} \not\models_{DB} \Psi(\boldsymbol{v})$

Steps of proof:

1. Choose $\tilde{\Psi}(\boldsymbol{v}) \in \text{ex}(pot\_sec(\mathcal{C}))$ arbitrarily
2. Show: There is a DB-Interpretation $\mathcal{I}^*$ with
   - $\mathcal{I}^* \models_M db_{f_s}$
   - $\mathcal{I}^* \models_M db_r$
   - $\mathcal{I}^* \models_M prior_{SC_R}$
   - $\mathcal{I}^* \not\models_M \tilde{\Psi}(\boldsymbol{v})$

Proof of Correctness (1)

About the structure of correct fragmentations

- Consider: $\tilde{\Psi}(\boldsymbol{v}) \in \mathrm{ex}(pot\_sec(\mathcal{C}))$ with $\boldsymbol{v} = (v_{i_1}, \ldots, v_{i_\ell})$
- Hence: $\tilde{\Psi}(\boldsymbol{X}) \in pot\_sec(\mathcal{C})$ with $\boldsymbol{X} = (X_{i_1}, \ldots, X_{i_\ell})$
- Moreover: $c = \{a_{i_1}, \ldots, a_{i_\ell}\} \in \mathcal{C}$
- Fragmentation $\mathcal{F}$ is correct w.r.t. $\mathcal{C}$
  - Accordingly: $c = \{a_{i_1}, \ldots, a_{i_\ell}\} \not\subseteq A_{F_s}$
  - Reformulated: There is $m \in \{i_1, \ldots, i_\ell\}$ s.t. $a_m \notin A_{F_s}$
- Hence: $m \notin \{1, \ldots, k\}$ and $m \in \{k+1, \ldots, n\}$

Proof of Correctness (1) – Visually Revisited

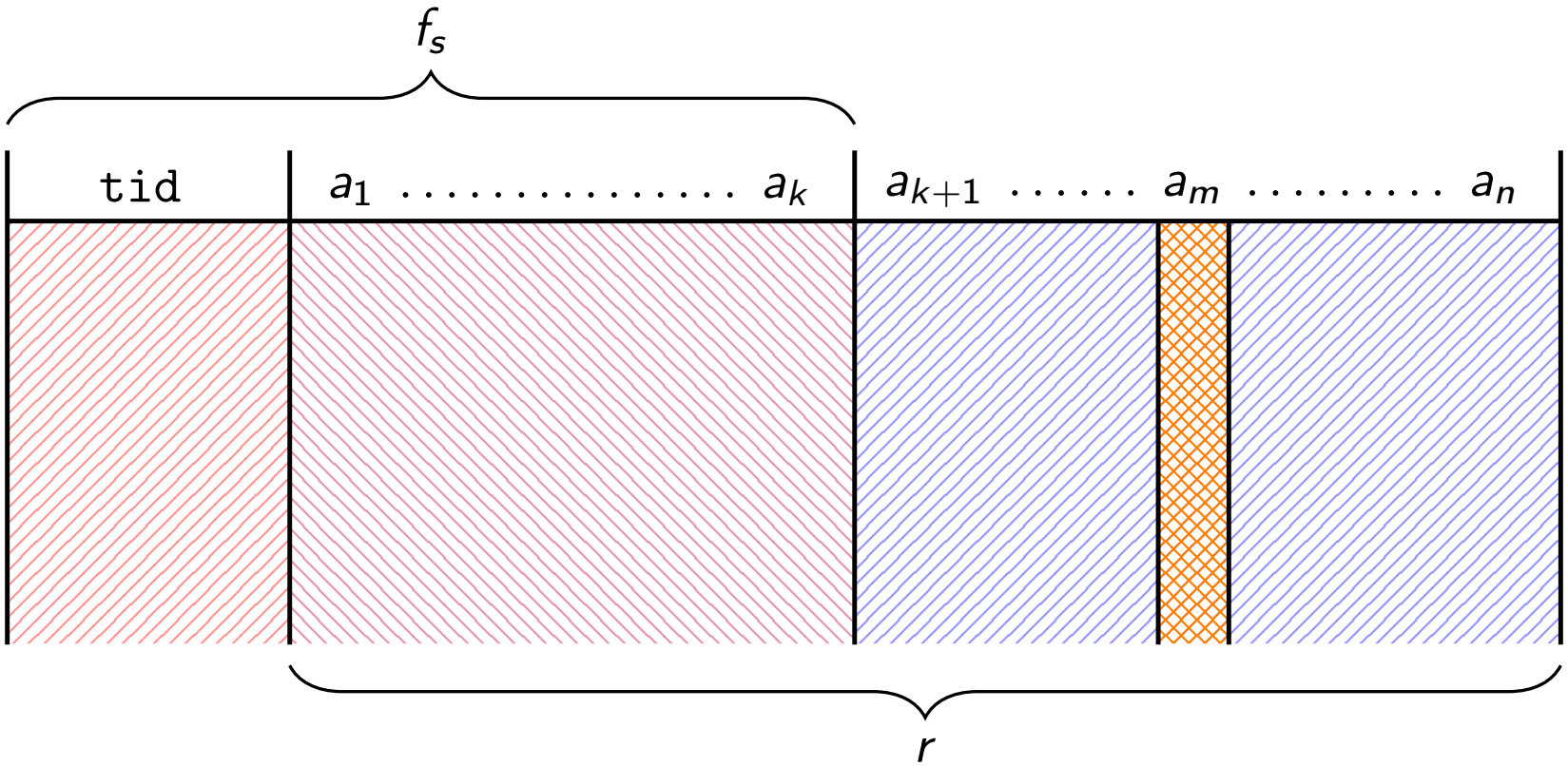


Figure: Properties $m \notin \{1, \ldots, k\}$ and $m \in \{k+1, \ldots, n\}$

Proof of Correctness (2)

First part of construction of $\mathcal{I}^*$:

$$\mathcal{I}^*(F_s) := \{ (\nu[a_{\mathtt{tid}}], \nu[a_1], \ldots, \nu[a_k]) \mid \nu \in f_s \}$$

Obviously $\mathcal{I}^* \models_M db_{f_s}$ because of

- $\mathcal{I}^* \models_M db^+_{f_s}$
- $\mathcal{I}^* \models_M db^-_{f_s}$
- $\mathcal{I}^* \models_M (a_{\mathtt{tid}} \to \{a_1, \ldots, a_k\})$

Proof of Correctness (3)

Continuing the construction of $\mathcal{I}^*$:

$$\mathcal{I}^*(R) := \{ (\mu[a_1], \ldots, \varphi_m(\mu[a_m]), \ldots, \mu[a_n]) \mid \mu \in r \}$$

$\varphi_m : \mathcal{U}_m \to \mathcal{U} \setminus \{v_m\}$ is an **injective** function with

- $\mathcal{U}_m := \{ \mu[a_m] \mid \mu \in r \}$
- $\mathcal{U}$ is the infinite universe of $\mathcal{I}^*$
- v_m is a value of $\boldsymbol{v} = (v_{i_1}, \ldots, v_{i_\ell})$

φ_m can always be constructed because of $||\mathcal{U} \setminus \{v_m\}|| > ||\mathcal{U}_m||$

Proof of Correctness (4)

First part of proving $\mathcal{I}^* \models_M db_r$: Show that $\mathcal{I}^*$ satisfies

$$(\forall X_1) \ldots (\forall X_k) \, [\, (\exists X_{\mathtt{tid}}) \, F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k) \;\Leftrightarrow$$
$$(\exists X_{k+1}) \ldots (\exists X_n) \, R(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n) \,]$$

To prove the if-part, assume:

$\mathcal{I}^* \models_M (\exists X_{\mathtt{tid}}) \, F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k)$ under $(X_1/u_1), \ldots, (X_k/u_k)$

- Hence: There is $(w_{\mathtt{tid}}, u_1, \ldots, u_k) \in \mathcal{I}^*(F_s)$
- Implies: $\nu \in f_s$ with $\nu[a_j] = u_j$ for $1 \leq j \leq k$
- By fragmentation: $\mu \in r$ with $\mu[a_j] = \nu[a_j]$ for $1 \leq j \leq k$
- As $m \notin \{1, \ldots, k\}$: $(u_1, \ldots, u_k, w_{k+1}, \ldots, w_n) \in \mathcal{I}^*(R)$

Only-if-part: Apply argumentation backwards!

Proof of Correctness (5) – Preparing Step

Second part of proving $\mathcal{I}^* \models_M db_r$: Show that $\mathcal{I}^*$ satisfies

$$(\forall X_1) \ldots (\forall X_k) [(\exists X_{\mathtt{tid}}) (\exists X'_{\mathtt{tid}}) [F_s(X_{\mathtt{tid}}, X_1, \ldots, X_k) \wedge F_s(X'_{\mathtt{tid}}, X_1, \ldots, X_k) \wedge (X_{\mathtt{tid}} \neq X'_{\mathtt{tid}})] \Rightarrow$$

$$(\exists X_{k+1}) \ldots (\exists X_n) (\exists X'_{k+1}) \ldots (\exists X'_n) [R(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n) \wedge R(X_1, \ldots, X_k, X'_{k+1}, \ldots, X'_n) \wedge \bigvee_{j=k+1}^{n} (X_j \neq X'_j)]]$$

Proof of Correctness (5)

Assume: $\mathcal{I}^* \models_M$ premise under $(X_1/u_1), \ldots, (X_k/u_k)$

- Hence, with $w_{\mathtt{tid}} \neq w'_{\mathtt{tid}}$
  - $(w_{\mathtt{tid}}, u_1, \ldots, u_k) \in \mathcal{I}^*(F_s)$
  - $(w'_{\mathtt{tid}}, u_1, \ldots, u_k) \in \mathcal{I}^*(F_s)$
- Implies: $\nu, \nu' \in f_s$ with $\nu[a_j] = \nu'[a_j] = u_j$ for $1 \leq j \leq k$
- By T-IDs: $\mu, \mu' \in r$ with $\mu[a_j] = \mu'[a_j] = u_j$ for $1 \leq j \leq k$
- No duplicates in $r \rightarrow \mu[a_p] \neq \mu'[a_p]$ for a $p \in \{k+1, \ldots, n\}$
- Accordingly
  - $(u_1, \ldots, u_k, w_{k+1}, \ldots, w_n) \in \mathcal{I}^*(R)$
  - $(u_1, \ldots, u_k, w'_{k+1}, \ldots, w'_n) \in \mathcal{I}^*(R)$
  - If $p \neq m$: Obviously $w_p \neq w'_p$
  - If $p = m$: $w_m \neq w'_m$ because φ_m is injective

Proof of Correctness (6)

To prove $\mathcal{I}^* \models_M prior_{SC_R}$: Construct temp. DB-Interpretation

$$\mathcal{I}_t(R) := \{ (\mu[a_1], \ldots, \mu[a_m], \ldots, \mu[a_n]) \mid \mu \in r \}$$

Obviously: $\mathcal{I}_t \models_M prior_{SC_R}$

About a DB-Interpretation $\mathcal{I}$ satisfying $prior_{SC_R}$

- Specific combinations of values in tuples not neccessary
- Only equalities and diversities in each column important

Between $\mathcal{I}_t$ and $\mathcal{I}^*$ holds:
$(u_1, \ldots, u_m, \ldots, u_n) \in \mathcal{I}_t(R)$ iff $(u_1, \ldots, \varphi_m(u_m), \ldots, u_n) \in \mathcal{I}^*(R)$

By injectivity: $u'_m = u''_m \quad$ iff $\quad \varphi_m(u'_m) = \varphi_m(u''_m)$

Proof of Correctness (7)

Last step to prove: $\mathcal{I}^* \not\models_M \tilde{\Psi}(\boldsymbol{v})$ with $\boldsymbol{v} = (v_{i_1}, \ldots, v_{i_\ell})$

$\mathcal{I}^* \models_M \tilde{\Psi}(\boldsymbol{v}) \Leftrightarrow$

- There is $(u_1, \ldots, u_m, \ldots, u_{|A_R|}) \in \mathcal{I}^*(R)$ with
- $u_j = v_j$ for **all** $j \in \{i_1, \ldots, i_\ell\}$:

This does not hold

- For all $(u_1, \ldots, u_m, \ldots, u_{|A_R|}) \in \mathcal{I}^*(R)$: $\varphi_m(\,\cdot\,) = u_m$
- $\varphi_m : \mathcal{U}_m \to \mathcal{U} \setminus \{v_m\}$
- $m \in \{i_1, \ldots, i_\ell\}$

q.e.d.

That's all...

Thank you for your attention!