technische universität dortmund

On the Inference-Proofness of Database Fragmentation Satisfying Confidentiality Constraints

Joachim Biskup **Marcel Preuß** Lena Wiese

Information Systems and Security (ISSI)

Technische Universität Dortmund, Germany

October 28, 2011

Table of Contents

Confidentiality by Fragmentation

Achieving Confidentiality by Breaking Associations

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

Often: Only associations between pieces of information sensitive

Example: Situation in a hospital

- List of illnesses cured $\rightsquigarrow$ Not sensitive
- List of patients $\rightsquigarrow$ Not really sensitive
- Association: Patient and his illness $\rightarrow$ Very sensitive

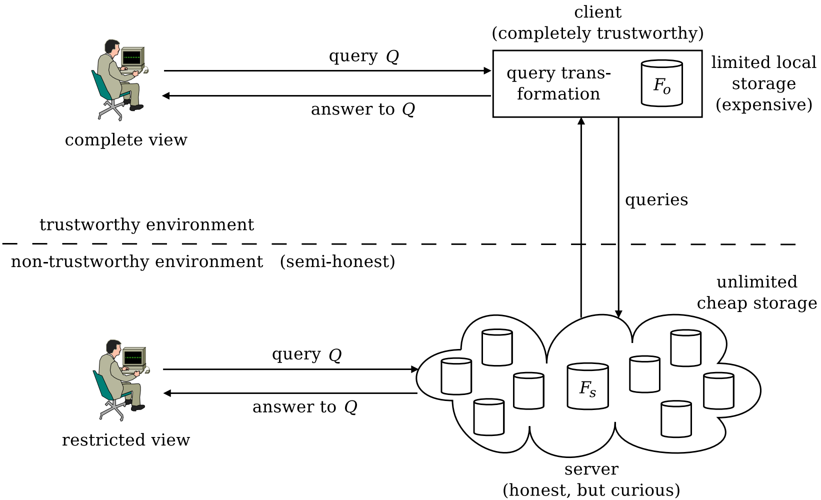
Goal: Confidentiality by breaking sensitive associations

Context of our contribution

Existing approach: Confidentiality by vertical fragmentation (by Samarati, Foresti, et al.)

- Formal framework of fragmentation
- Formal declaration of confidentiality requirements
- Efficient computation of fragmented instances
- Answering queries over fragmented databases
- No formal proof of inference-proofness

Scenario for Working with Fragmented Databases



technische universität dortmund

Fragmentation Compliant with Scenario

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

- 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 relational schema with $A_{F_i} \subseteq A_R$
  - Each attribute of A_R either in A_{F_o} or A_{F_s}
- 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
  - External storage of instance f_s
- Tuples belonging together have a unique Tuple-ID in common

technische universität dortmund

Example: Instance containing sensitive associations

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

technische universität dortmund

Example: Possible Fragmentation with Tuple-IDs

F_o (local)	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 (external)	tid	ZIP	Illness	Doctor
	1	94142	Hypertension	White
	2	94141	Obesity	Warren
	3	94142	Hypertension	White
	4	94139	Obesity	Warren

Formal Declaration of Confidentiality Requirements

How to declare confidentiality requirements formally?

Confidentiality Constraint c over $\langle R|A_R|SC_R\rangle$: Attributes $c \subseteq A_R$

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
- $\mathcal{F}$ correct w.r.t. $\mathcal{C} \quad \Leftrightarrow \quad c \not\subseteq A_{F_s}$ holds for all $c \in \mathcal{C}$

Example: Correct Fragmentation

F_o (local)	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 (external)	tid	ZIP	Illness	Doctor
	1	94142	Hypertension	White
	2	94141	Obesity	Warren
	3	94142	Hypertension	White
	4	94139	Obesity	Warren

is correct w.r.t.

$$\mathcal{C} = \{ \quad c_1 = \{\texttt{SSN}\}, \quad c_3 = \{\texttt{Name}, \texttt{Illness}\}, \quad c_2 = \{\texttt{Name}, \texttt{DoB}\}, \quad c_4 = \{\texttt{DoB}, \texttt{ZIP}, \texttt{Illness}\} \quad \}$$

Inference-Proofness of Fragmentation

Approach to Show Inference-Proofness

How to analyse inference-proofness?

- Controlled Query Evaluation (CQE)
  is known to be inference-proof
- Logic-oriented modelling of fragmentation
  within CQE-Framework
  from the point of view of an attacker
- Formal proof within logic-oriented framework

Modelling the Positive Knowledge of f_s

Suppose: Attacker knows

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

Explicit positive knowlegde from attacker's point of view

$$\{\, F_s(\, \nu[a_{\texttt{tid}}], \nu[a_1], \ldots, \nu[a_k]\,) \mid \nu \in f_s \,\}$$

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

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

Negative Knowledge Resulting from Completeness

Problem: An attacker knows even more about f_s

- Instances supposed to be complete
- By CWA: Every constant combination not in f_s is invalid
  $\rightarrow$ Knowledge of the kind $\neg F_s\,(v_{\texttt{tid}}, v_1, \ldots, v_k)$
- Problem: Infinite set of constant symbols
- Express negative knowledge by Completeness Sentence

technische universität dortmund

The Knowledge About the Hidden Instance r

Suppose: Attacker knows the process of fragmentation including

- Outsourced 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 original instance r is built

Knowledge resulting from relationship between f_s and r

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

This knowledge must be expressed as a logic formula!

Confidentiality Constraints in the CQE-Framework

Confidentiality constraints modelled as potential secrets

- Potential secret Ψ in CQE-framework:
  - Ψ is a logic sentence
  - If Ψ is true in instance: User must *not* get to know this
  - Otherwise: User may know that Ψ is false in instance
- Consider confidentiality constraint $c_i = \{a_{i_1}, \ldots, a_{i_\ell}\}$
- Protect *all* constant combinations possible for $a_{i_1}, \ldots, a_{i_\ell}$

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

technische universität dortmund

About A-Priori Knowledge

Preliminary result:

- Logic-oriented view on fragmentation
- Attacker's a priori knowledge neglected so far

But: A priori knowledge of crucial importance

- No inference-proofness under general a priori knowledge
- Here: Inference-proofness under EGDs/TGDs which are
  - Unirelational
  - Typed
  - Without Constants

About Unirelational Typed EGDs/TGDs

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

- Equality Generating Dependencies (EGDs) (e.g., FDs)
  - Presence of some tuples in r implies:
    Certain components of these tuples are equal
- Tuple Generating Dependencies (TGDs) (e.g., JDs, INDs)
  - Presence of some tuples in r implies:
    Presence of certain other tuples in r
- Typed: Assignment of variables to column positions

Main result: Inference-Proofness

To be shown: For each potential secret Ψ

$$\left.\begin{array}{l}\text{Knowledge about outsourced instance } f_s \\ \text{Knowledge about hidden instance } r \\ \text{A priori knowledge: Unirel. typed EGDs/TGDs}\end{array}\right\} \quad \not\models \quad \Psi$$

Sketch of proof:

1. Choose any potential secret $\tilde{\Psi}$
2. Construct an interpretation $\mathcal{I}^*$ with
   - $\mathcal{I}^* \models_M$ Knowledge about outsourced instance f_s
   - $\mathcal{I}^* \models_M$ Knowledge about hidden instance r
   - $\mathcal{I}^* \models_M$ A priori knowledge: Unirel. typed EGDs/TGDs
   - $\mathcal{I}^* \not\models_M \tilde{\Psi}$

Conclusion and Future Work

Conclusion and Future Work

What has been achieved?

- Existing approach to confidentiality by fragmentation is
  - Modelled logic-orientedly in CQE-framework
  - Extended by attacker's a priori knowledge
- Within modelling: Formal proof of inference-proofness

What might be done in future?

- Extending feasible a priori knowledge
- Analysing other approaches to confidentiality by fragmentation
- Hybrid fragmentation: Vertical + Horizontal

technische universität
dortmund

That's all...

Thank you for your attention!