technische universität dortmund

Database Fragmentation with Encryption: Under Which Semantic Constraints and A Priori Knowledge Can Two Keep a Secret?

Joachim Biskup Marcel Preuß

Information Systems and Security (ISSI)

Technische Universität Dortmund, Germany

March 11, 2013

technische universität dortmund

Table of Contents

Confidentiality by Fragmentation

technische universität dortmund

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 Aggarwal, Bawa, et al.)

- Formal framework of fragmentation (More or less)
- Formal declaration of confidentiality requirements
- Efficient computation of fragmented instances
- Answering queries over fragmented databases
- **No formal proof** of inference-proofness

Towards an Approach to Fragmentation

Assumptions: Underlying client-server framework

- Two servers, both honest, but curious
- No cooperation between servers
- Each server stores exactly one of two fragments
- Attacker has access to at most one server
- No persistent local storage
  - All data must be stored externally
  - Client only processes queries
- Authorized user has access to both servers (via client)

Assumptions About the Encryption Function

Approach employs encryption within fragmentation

Encryption function $Enc : \mathcal{U} \times \mathcal{U} \to \mathcal{U}$ satisfies group properties

- Each value of $\mathcal{U}$ can be a
  - Plaintext v
  - Cryptographic key κ
  - Ciphertext e
- Given an arbitrary pair of two values $\in \{v, \kappa, e\}$
  The missing value $\in \{v, \kappa, e\}$ can be determined s.t.
  $Enc(v, \kappa) = e$ holds
- Decryption function: $Dec(e, \kappa) = v$ iff $Enc(v, \kappa) = e$

Fragmentation Compliant with Assumptions

Fragmentation $(\mathcal{F}, \mathcal{E})$ of instance r over schema $\langle R|A_R|SC_R\rangle$

- On **schema** level
  - Distinguished attribute $a_{\mathtt{tid}} \notin A_R$ for tuple identifiers (TIDs)
  - Set of "encrypted attributes" $\mathcal{E} \subseteq A_R$
  - Set of fragments $\mathcal{F} = \{\langle F_1|A_{F_1}|SC_{F_1}\rangle, \langle F_2|A_{F_2}|SC_{F_2}\rangle\}$
    - $A_{F_i} := \{a_{\mathtt{tid}}\} \cup \bar{A}_{F_i}$ with $\bar{A}_{F_i} \subseteq A_R$
    - $SC_{F_i} := \{a_{\mathtt{tid}} \rightarrow \bar{A}_{F_i}\}$ (Functional dependency)
    - $\bar{A}_{F_1} \cup \bar{A}_{F_2} = A_R$ and $\bar{A}_{F_1} \cap \bar{A}_{F_2} = \mathcal{E}$
- On **instance** level
  - Instances f_1 over $\langle F_1|A_{F_1}|SC_{F_1}\rangle$ and f_2 over $\langle F_2|A_{F_2}|SC_{F_2}\rangle$
  - For each $\mu \in r$: exactly one $\nu_1 \in f_1$, exactly one $\nu_2 \in f_2$ with
    - $\nu_1[a_{\mathtt{tid}}] = \nu_2[a_{\mathtt{tid}}] = v_\mu$ s.t. v_μ is globally unique
    - $\nu_i[a] := \mu[a]$ for each $a \in \left(\bar{A}_{F_i} \setminus \mathcal{E}\right)$, $i \in \{1, 2\}$
    - $\nu_1[a] := Enc\,(\mu[a], \kappa)$ and $\nu_2[a] := \kappa$ for each $a \in \mathcal{E}$ s.t. κ is random but globally unique f.e. $\mu \in r$, $a \in \mathcal{E}$

Fragmentation of Example Instance

R	SSN	Name	Illness	HurtBy	Doctor
	1234	Hellmann	Borderline	Hellmann	White
	2345	Dooley	Laceration	McKinley	Warren
	3456	McKinley	Laceration	Dooley	Warren
	3456	McKinley	Concussion	Dooley	Warren

$$\Downarrow$$

F_1	<u>tid</u>	SSN	Name	HurtBy	Doctor
	1	e_S^1	Hellmann	e_H^1	White
	2	e_S^2	Dooley	e_H^2	Warren
	3	e_S^3	McKinley	e_H^3	Warren
	4	e_S^4	McKinley	e_H^4	Warren

F_2	<u>tid</u>	SSN	HurtBy	Illness
	1	κ_S^1	κ_H^1	Borderline
	2	κ_S^2	κ_H^2	Laceration
	3	κ_S^3	κ_H^3	Laceration
	4	κ_S^4	κ_H^4	Concussion

SSN and **HurtBy** are "encrypted attributes"

Convention from now on

Consider: Rearrangement of columns of instances r, f_1, f_2

Suppose: $A_R = \{a_1, \ldots, a_h, a_{h+1}, \ldots, a_k, a_{k+1}, \ldots, a_n\}$ s.t.

	$A_{F_i} \setminus A_R$	$(A_{F_1} \setminus \mathcal{E}) \cap A_R$	$\mathcal{E} \cap A_{F_i} \cap A_R$	$(A_{F_2} \setminus \mathcal{E}) \cap A_R$
A_R		$a_1, \ldots, a_h$	$a_{h+1}, \ldots, a_k$	$a_{k+1}, \ldots, a_n$
A_{F_1}	$a_{\mathtt{tid}}$	$a_1, \ldots, a_h$	$a_{h+1}, \ldots, a_k$	
A_{F_2}	$a_{\mathtt{tid}}$		$a_{h+1}, \ldots, a_k$	$a_{k+1}, \ldots, a_n$

Attention: For $j \in \{h+1, \ldots, k\}$: Same attributes, different values

- Tuple $\mu \in r$: $\mu[a_j]$ is a plaintext value
- Tuple $\nu_1 \in f_1$: $\nu_1[a_j]$ is a ciphertext value
- Tuple $\nu_2 \in f_2$: $\nu_2[a_j]$ is a cryptographic key

Reconstructability of Original Instance r

Given: Fragment-instances f_1 and f_2 of original instance r

For $\nu_1 \in f_1$, $\nu_2 \in f_2$ with $\nu_1[a_{\texttt{tid}}] = \nu_2[a_{\texttt{tid}}]$:

$$
\begin{aligned}
\nu_1 \diamond \nu_2 = (\ & \nu_1[a_1], \ldots, \nu_1[a_h], \\
& Dec(\nu_1[a_{h+1}], \nu_2[a_{h+1}]), \ldots, Dec(\nu_1[a_k], \nu_2[a_k]), \\
& \nu_2[a_{k+1}], \ldots, \nu_2[a_n] \quad)
\end{aligned}
$$

By fragmentation: $\nu_1 \diamond \nu_2 \in r$

For $\nu_1 \in f_1$, $\nu_2 \in f_2$ with $\nu_1[a_{\texttt{tid}}] \neq \nu_2[a_{\texttt{tid}}]$:
$\nu_1 \diamond \nu_2$ is undefined

Formal Declaration of Confidentiality Requirements

How to declare confidentiality requirements?

Syntax: Confidentiality Constraint c over $\langle R|A_R|SC_R\rangle$:
Non-empty subset $c \subseteq A_R$ of attributes

Semantics: Confidentiality of fragmentation

- Let $\mathcal{C}$ be a set of Confidentiality Constraints
- Fragmentation $(\mathcal{F}, \mathcal{E})$ is confidential w.r.t. $\mathcal{C}$ $\Leftrightarrow$
  For $i \in \{1, 2\}$: $c \nsubseteq (A_{F_i} \setminus \mathcal{E})$ for all $c \in \mathcal{C}$

Confidential Fragmentation of Example Instance

R	SSN	Name	Illness	HurtBy	Doctor
	1234	Hellmann	Borderline	Hellmann	White
	2345	Dooley	Laceration	McKinley	Warren
	3456	McKinley	Laceration	Dooley	Warren
	3456	McKinley	Concussion	Dooley	Warren

F_1	<u>tid</u>	SSN	Name	HurtBy	Doctor
	1	e_S^1	Hellmann	e_H^1	White
	2	e_S^2	Dooley	e_H^2	Warren
	3	e_S^3	McKinley	e_H^3	Warren
	4	e_S^4	McKinley	e_H^4	Warren

F_2	<u>tid</u>	SSN	HurtBy	Illness
	1	κ_S^1	κ_H^1	Borderline
	2	κ_S^2	κ_H^2	Laceration
	3	κ_S^3	κ_H^3	Laceration
	4	κ_S^4	κ_H^4	Concussion

is confidential w.r.t.

$$\mathcal{C} = \{ \quad c_1 = \{\mathtt{SSN}\}, \quad c_3 = \{\mathtt{Name}, \mathtt{HurtBy}\}, \quad c_2 = \{\mathtt{Name}, \mathtt{Illness}\}, \quad c_4 = \{\mathtt{Illness}, \mathtt{HurtBy}\} \quad \}$$

Inference-Proofness of Fragmentation

Approach to Show Inference-Proofness

How to analyze inference-proofness?

- Controlled Interaction Execution (CIE) is known to be inference-proof
- Logic-oriented modelling of fragmentation within CIE-Framework from attacker's point of view
- Formal proof within logic-oriented framework

Construction of an Appropriate Logic: Syntax

Language $\mathscr{L}$: First-order logic with equality

- Set $\mathcal{P}$ of predicate symbols
  - F_1 with arity $k+1 = |A_{F_1}|$
  - F_2 with arity $n-h+1 = |A_{F_2}|$
  - R with arity $n = |A_R|$
- Distinguished binary predicate symbol $\equiv$
- A term of an atomic formula can be a
  - Binary function symbol E, D
  - Constant symbol of fixed infinite domain *Dom*
  - Variable of infinite set $Var := \{X_1, X_2, \ldots, Y_1, Y_2, \ldots\}$

Construction of an Appropriate Logic: Semantics

Interpretation $\mathcal{I}$ for $\mathscr{L}$ is a DB-Interpretation iff

- Universe $\mathcal{U} := \mathcal{I}(Dom) = Dom$
- $\mathcal{I}(v) = v$ for all $v \in Dom$
- $\mathcal{I}(E)(v, \kappa) = e$ iff $Enc(v, \kappa) = e$
- $\mathcal{I}(D)(e, \kappa) = v$ iff $Dec(e, \kappa) = v$
- $P \in \mathcal{P}$ with arity m is interpreted by finite set $\mathcal{I}(P) \subset \mathcal{U}^m$
- $\mathcal{I}(\equiv) = \{ (v, v) \mid v \in \mathcal{U} \}$

Complete instances r, f_1 and f_2 induce DB-Interpretation $\mathcal{I}_r$

- $(v_1, \ldots, v_n) \in \mathcal{I}_r(R)$ iff $(v_1, \ldots, v_n) \in r$
- Analogously for $\mathcal{I}_r(F_1)$, $\mathcal{I}_r(F_2)$ induced by fragments f_1, f_2 of r

Satisfaction and Implication Based on DB-Interpretation

Satisfaction of sentences (closed formulas) of $\mathscr{L}$

- Notation of satisfaction
  - Consider: DB-Interpretation $\mathcal{I}$, set of sentences $\mathcal{S} \subset \mathscr{L}$
  - $\mathcal{I}$ satisfies $\mathcal{S}$ written as $\mathcal{I} \models_M \mathcal{S}$
- Semantics of satisfaction: Same as in usual first-order logic

Implication based on DB-Interpretation

- Notation: $\mathcal{S} \subset \mathscr{L}$ implies $\Phi \in \mathscr{L}$ written as $\mathcal{S} \models_{DB} \Phi$
- Semantics: $\mathcal{S} \models_{DB} \Phi$ iff
  For each DB-Interpretation $\mathcal{I}$: If $\mathcal{I} \models_M \mathcal{S}$ then $\mathcal{I} \models_M \Phi$

Modelling the Positive Knowledge of f_1

Suppose: Attacker knows

- Outsourced fragment instance f_1
- Fragment $\langle F_1 | A_{F_1} | SC_{F_1} \rangle$ with $A_{F_1} = \{a_{\mathtt{tid}}, a_1, \ldots, a_k\}$

Attacker's explicit positive knowlegde of f_1

- $db^+_{f_1} := \{F_1(\nu[a_{\mathtt{tid}}], \nu[a_1], \ldots, \nu[a_k]) \mid \nu \in f_1\}$
- Functional dependency $a_{\mathtt{tid}} \rightarrow \{a_1, \ldots, a_k\} \in SC_{F_1}$

Negative Knowledge Resulting from Completeness

Problem: An attacker knows even more about f_1

- Instances r, f_1 and f_2 are supposed to be complete
- Every constant combination not in f_1 is invalid by CWA
  $\rightarrow$ Knowledge of the kind $\neg F_1\,(v_{\mathtt{tid}}, v_1, \ldots, v_k)$
- Problem: Infinite Domain $\rightarrow$ Not explicitly enumerable
- Bright idea: Use Completeness-Sentence to model CWA

Construction of Completeness Sentence: Example

F_1	tid	SSN	Name	HurtBy	Doctor
	1	e_S^1	Hellmann	e_H^1	White
	2	e_S^2	Dooley	e_H^2	Warren
	3	e_S^3	McKinley	e_H^3	Warren
	4	e_S^4	McKinley	e_H^4	Warren

Completeness sentence resulting from f_1:

$$
\begin{array}{l}
(\forall X_t)(\forall X_S)(\forall X_N)(\forall X_H)(\forall X_D)\ [\\
(X_t \equiv 1 \wedge X_S \equiv e_S^1 \wedge X_N \equiv \text{Hellmann} \wedge X_H \equiv e_H^1 \wedge X_D \equiv \text{White}) \vee\\
(X_t \equiv 2 \wedge X_S \equiv e_S^2 \wedge X_N \equiv \text{Dooley} \wedge X_H \equiv e_H^2 \wedge X_D \equiv \text{Warren}) \vee\\
(X_t \equiv 3 \wedge X_S \equiv e_S^3 \wedge X_N \equiv \text{McKinley} \wedge X_H \equiv e_H^3 \wedge X_D \equiv \text{Warren}) \vee\\
(X_t \equiv 4 \wedge X_S \equiv e_S^4 \wedge X_N \equiv \text{McKinley} \wedge X_H \equiv e_H^4 \wedge X_D \equiv \text{Warren}) \vee\\
\neg F_1(X_t, X_S, X_N, X_H, X_D)\ ]
\end{array}
$$

Modelling the Negative Knowledge of f_1

Completeness sentence for running example:

$$
\begin{array}{l}
(\forall X_t)(\forall X_S)(\forall X_N)(\forall X_H)(\forall X_D)\ [\\
(X_t \equiv 1 \land X_S \equiv e_S^1 \land X_N \equiv \text{Hellmann} \land X_H \equiv e_H^1 \land X_D \equiv \text{White}) \lor\\
(X_t \equiv 2 \land X_S \equiv e_S^2 \land X_N \equiv \text{Dooley} \land X_H \equiv e_H^2 \land X_D \equiv \text{Warren}) \lor\\
(X_t \equiv 3 \land X_S \equiv e_S^3 \land X_N \equiv \text{McKinley} \land X_H \equiv e_H^3 \land X_D \equiv \text{Warren}) \lor\\
(X_t \equiv 4 \land X_S \equiv e_S^4 \land X_N \equiv \text{McKinley} \land X_H \equiv e_H^4 \land X_D \equiv \text{Warren}) \lor\\
\neg F_1(X_t, X_S, X_N, X_H, X_D) \quad]
\end{array}
$$

Construction of Completeness Sentence of $db_{f_1}^-$ in general:

$$
(\forall X_{\texttt{tid}}) \ldots (\forall X_k) \left[\bigvee_{\nu \in f_1} \left(\bigwedge_{a_j \in A_{F_1}} (X_j \equiv \nu[a_j]) \right) \lor \neg F_1(X_{\texttt{tid}}, X_1, \ldots, X_k) \right]
$$

Final Logic-Oriented View on f_1

Summing up: A logic-oriented view on f_1 is modelled by

$$db_{f_1} := db_{f_1}^{+} \cup db_{f_1}^{-} \cup \{a_{\mathtt{tid}} \rightarrow \{a_1, \ldots, a_k\}\}$$

But: Attacker is curious about original instance r
(or f_2, respectively)

Attacker's Knowledge About r and f_2 (1)

Suppose: Attacker knows

- Schema $\langle R|A_R|SC_R\rangle$ over which original instance r is built
- Process of fragmentation (algorithm)
- Computed fragmentation $\mathcal{F} = \{\langle F_1|A_{F_1}|SC_{F_1}\rangle, \langle F_2|A_{F_2}|SC_{F_2}\rangle\}$

Suppose: Attacker has **no** access to

- Original instance r (not materialized at all)
- Fragment instance f_2 (hosted by "other" server)

Suppose: Attacker is curious about r (or f_2, respectively)

Attacker's Knowledge About r and f_2 (2)

Attacker's deductions: For each $\nu_1 \in f_1$

- Tuple $\nu_2 \in f_2$ with $\nu_2[a_{\texttt{tid}}] = \nu_1[a_{\texttt{tid}}]$ exists
- Tuple $\mu \in r$ with $\nu_1 \diamond \nu_2 = \mu$ exists

Knowledge expressed as a sentence of db_R:

$$
\begin{aligned}
&(\forall X_{\texttt{tid}})\,(\forall X_1)\ldots(\forall X_h)\,(\forall X_{h+1})\ldots(\forall X_k)\,\Big[\\
&\quad F_1\,(X_{\texttt{tid}}, X_1, \ldots, X_h, X_{h+1}, \ldots, X_k)\\
&\qquad \Rightarrow\\
&\quad (\exists Y_{h+1})\ldots(\exists Y_k)\,(\exists Z_{k+1})\ldots(\exists Z_n)\,\Big[\\
&\qquad F_2\,(X_{\texttt{tid}}, Y_{h+1}, \ldots, Y_k, Z_{k+1}, \ldots, Z_n) \wedge\\
&\qquad R\,(X_1, \ldots, X_h, D\,(X_{h+1}, Y_{h+1}), \ldots, D\,(X_k, Y_k), Z_{k+1}, \ldots, Z_n)\Big]\Big]
\end{aligned}
$$

Attacker's Knowledge About r and f_2 (3)

The equivalence does **not** hold!

Supposed fragmentation with "encrypted attribute" a_2:

R	a_1	a_2	a_3
	v_1	v_2	v_3
	v_1'	v_2	v_3

F_1	$a_{\mathtt{tid}}$	a_1	a_2
	1	v_1	c_2
	2	v_1'	c_2'

F_2	$a_{\mathtt{tid}}$	a_2	a_3
	1	κ_2	v_3
	2	κ_2'	v_3

Implication possible under equivalence:

$$\left[\, F_2(1,\kappa_2,v_3) \wedge R(v_1', \overbrace{D(\square,\kappa_2)}^{=v_2}, v_3) \,\right] \Rightarrow F_1(1, v_1', \square)$$

By properties of perfect encryption: $D(\square,\kappa_2) = v_2$ iff $\square = c_2$
$\rightarrow$ Tuple $(1, v_1', c_2) \in f_1$ $\lightning$

Attacker's Knowledge About r and f_2 (4)

Attacker's deductions: Tuple $\nu_2 \in f_2$ can *only* exist if

- Tuple $\nu_1 \in f_1$ with $\nu_1[a_{\mathtt{tid}}] = \nu_2[a_{\mathtt{tid}}]$ exists
- Tuple $\mu \in r$ with $\nu_1 \diamond \nu_2 = \mu$ exists

Knowledge expressed as a sentence of db_R:

$$
\begin{aligned}
&(\forall X_{\mathtt{tid}})\,(\forall X_{h+1}) \dots (\forall X_k)\,(\forall X_{k+1}) \dots (\forall X_n)\,\Big[\\
&\quad F_2\,(X_{\mathtt{tid}}, X_{h+1}, \dots, X_k, X_{k+1}, \dots, X_n)\\
&\qquad \Rightarrow\\
&\quad (\exists Y_1) \dots (\exists Y_h)\,(\exists Z_{h+1}) \dots (\exists Z_k)\,\Big[\\
&\qquad F_1\,(X_{\mathtt{tid}}, Y_1, \dots, Y_h, Z_{h+1}, \dots, Z_k) \wedge\\
&\qquad R\,(Y_1, \dots, Y_h, D\,(Z_{h+1}, X_{h+1}), \dots, D\,(Z_k, X_k), X_{k+1}, \dots, X_n)\Big]\Big]
\end{aligned}
$$

Attacker's Knowledge About r and f_2 (5)

The equivalence does **not** hold!

Supposed fragmentation with "encrypted attribute" a_2:

R	a_1	a_2	a_3
	v_1	v_2	v_3
	v_1	v_2	v_3'

F_1	$a_{\mathtt{tid}}$	a_1	a_2
	1	v_1	c_2
	2	v_1	c_2'

F_2	$a_{\mathtt{tid}}$	a_2	a_3
	1	κ_2	v_3
	2	κ_2'	v_3'

Implication possible under equivalence:

$$\left[\, F_1(1, v_1, c_2) \wedge R(v_1, \overbrace{D(c_2, \square)}^{= v_2}, v_3') \,\right] \Rightarrow F_2(1, \square, v_3')$$

By properties of perfect encryption: $D(c_2, \square) = v_2$ iff $\square = \kappa_2$
$\rightarrow$ Tuple $(1, \kappa_2, v_3') \in f_2$ ↯

Attacker's Knowledge About r and f_2 (6)

Attacker's deductions: Tuple $\mu \in r$ exists iff

- Tuples $\nu_1 \in f_1$ and $\nu_2 \in f_2$ with $\nu_1[a_{\mathtt{tid}}] = \nu_2[a_{\mathtt{tid}}]$ exist s.t.
- $\nu_1 \diamond \nu_2 = \mu$ holds

Knowledge expressed as a sentence of db_R:

$$
\begin{aligned}
&(\forall X_1)\ldots(\forall X_h)(\forall X_{h+1})\ldots(\forall X_k)(\forall X_{k+1})\ldots(\forall X_n)\Big[\\
&\quad R(X_1,\ldots,X_h,X_{h+1},\ldots,X_k,X_{k+1},\ldots,X_n)\\
&\qquad \Leftrightarrow\\
&\quad (\exists Z_{\mathtt{tid}})(\exists Y_{h+1})\ldots(\exists Y_k)\Big[\\
&\qquad F_2(Z_{\mathtt{tid}},Y_{h+1},\ldots,Y_k,X_{k+1},\ldots,X_n)\wedge\\
&\qquad F_1(Z_{\mathtt{tid}},X_1,\ldots,X_h,E(X_{h+1},Y_{h+1}),\ldots,E(X_k,Y_k))\Big]\Big]
\end{aligned}
$$

Here: Equivalence holds by fragmentation!

Attacker's Knowledge About r and f_2 (7)

Attacker's deductions: By fragmentation and tuple-IDs

- If different tuples $\nu_1, \nu_1' \in f_1$ are equal w.r.t. $(A_{F_1} \cap A_R) \setminus \mathcal{E}$, corresponding $\mu, \mu' \in r$ are equal w.r.t. $(A_{F_1} \cap A_R) \setminus \mathcal{E}$
- But: μ and μ' cannot be duplicates

Knowledge expressed as a sentence of db_R:

$$
\begin{aligned}
&(\forall X_{\mathtt{tid}})\,(\forall X'_{\mathtt{tid}})\,(\forall X_1)\ldots(\forall X_h)\,(\forall X_{h+1})\ldots(\forall X_k)\,(\forall X'_{h+1})\ldots(\forall X'_k)\,\Big[\\
&\quad\Big[F_1\,(X_{\mathtt{tid}}, X_1, \ldots, X_h, X_{h+1}, \ldots, X_k) \wedge\\
&\quad\; F_1\,\big(X'_{\mathtt{tid}}, X_1, \ldots, X_h, X'_{h+1}, \ldots, X'_k\big) \wedge (X_{\mathtt{tid}} \neq X'_{\mathtt{tid}})\Big]\\
&\qquad \Rightarrow\\
&\quad(\exists Y_{h+1})\ldots(\exists Y_n)\,(\exists Z_{h+1})\ldots(\exists Z_n)\,\Big[\\
&\qquad R\,(X_1, \ldots, X_h, Y_{h+1}, \ldots, Y_k, Y_{k+1}, \ldots, Y_n) \wedge\\
&\qquad R\,(X_1, \ldots, X_h, Z_{h+1}, \ldots, Z_k, Z_{k+1}, \ldots, Z_n) \wedge \textstyle\bigvee_{j=h+1}^{n} (Y_j \neq Z_j)\,\Big]\Big]
\end{aligned}
$$

Attacker's Knowledge About r and f_2 (8)

Attacker's deductions: By fragmentation and tuple-IDs

- If different tuples $\nu_2, \nu_2' \in f_2$ are equal w.r.t. $(A_{F_2} \cap A_R) \setminus \mathcal{E}$, corresponding $\mu, \mu' \in r$ are equal w.r.t. $(A_{F_2} \cap A_R) \setminus \mathcal{E}$
- But: μ and μ' cannot be duplicates

Knowledge expressed as a sentence of db_R:

$$
\begin{array}{l}
(\forall X_{\mathtt{tid}})\,(\forall X'_{\mathtt{tid}})\,(\forall X_{h+1}) \ldots (\forall X_k)\,\left(\forall X'_{h+1}\right) \ldots (\forall X'_k)\,(\forall X_{k+1}) \ldots (\forall X_n)\,\Big[\\
\quad \big[F_2\,(X_{\mathtt{tid}}, X_{h+1}, \ldots, X_k, X_{k+1}, \ldots, X_n) \wedge \\
\quad\; F_2\,\left(X'_{\mathtt{tid}}, X'_{h+1}, \ldots, X'_k, X_{k+1}, \ldots, X_n\right) \wedge (X_{\mathtt{tid}} \neq X'_{\mathtt{tid}})\big] \\
\qquad \Rightarrow \\
\quad (\exists Y_1) \ldots (\exists Y_k)\,(\exists Z_1) \ldots (\exists Z_k)\,\Big[\\
\qquad R\,(Y_1, \ldots, Y_h, Y_{h+1}, \ldots, Y_k, X_{k+1}, \ldots, X_n) \wedge \\
\qquad R\,(Z_1, \ldots, Z_h, Z_{h+1}, \ldots, Z_k, X_{k+1}, \ldots, X_n) \wedge \bigvee_{j=1}^{k} (Y_j \neq Z_j)\,\Big]\Big]
\end{array}
$$

Confidentiality Constraints in the CIE-Framework

Design choice: Confidentiality constraints as potential secrets

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

Bridging the Differences

From Confidentiality Constraints to Potential Secrets

- Consider a confidentiality constraint $c_i = \{a_{i_1}, \dots, a_{i_\ell}\}$
- Protect *all* constant combinations possible for $a_{i_1}, \dots, a_{i_\ell}$
  - Otherwise: Attacker can read secrets directly from $potsec(\mathcal{C})$
  - But: Leads to an infinite number of sentences (as $|Dom| = \infty$)
    $\rightarrow$ One potential secret per possible constant combination
- Use free variables $X_{i_1}, \dots, X_{i_\ell}$ to represent $a_{i_1}, \dots, a_{i_\ell}$

Modelling of Confidentiality Constraints

Consider: Confidentiality constraint $c_i \in \mathcal{C}$

- $c_i = \{a_{i_1}, \dots, a_{i_\ell}\} \subseteq \{a_1, \dots, a_n\} = A_R$
- $A_R \setminus c_i = \{a_{i_{\ell+1}}, \dots, a_{i_n}\}$

Construction of *potsec*$(\mathcal{C})$:

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

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

- $\boldsymbol{X_i} = (X_{i_1}, \dots, X_{i_\ell})$ is the vector of free variables of $\Psi_i(\boldsymbol{X_i})$

Expansion of the Confidentiality Policy

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

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

- Consider each $\boldsymbol{v_i} = (v_{i_1}, \ldots, v_{i_\ell}) \in Dom^\ell$
- Construct each sentence $\Psi_i(\boldsymbol{v_i})$

Expansion of $potsec(\mathcal{C})$:

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

The Impact of A-Priori Knowledge: Survey

Known now: Logic-oriented view on fragmentation

Until now: Attacker's a priori knowledge has been neglected

- Knowledge about the world in general
- Knowledge about semantic database constraints SC_R

Survey of the following results

- No inference-proofness under general a priori knowledge $\lightning$
- Inference-proofness under constrained a priori knowledge ✓

Goal: Construction of confidential fragmentation
Complying with a priori knowledge

The Impact of A Priori Knowledge: Example (1)

Attacker's view on r based on f_1:

R	SSN	Name	Illness	HurtBy	Doctor
	?	Hellmann	?	?	White
	?	Dooley	?	?	Warren
	?	McKinley	?	?	Warren
	?	McKinley	?	?	Warren

Suppose attacker knows a priori:
"All patients of psychiatrist `White` suffer from `Borderline`."

As a sentence of $\mathscr{L}$:
$(\forall X_S)(\forall X_N)(\forall X_I)(\forall X_H)\,[\,R(X_S, X_N, X_I, X_H, \texttt{White}) \Rightarrow (X_I \equiv \texttt{BLine})\,]$

Attacker's updated view on r violates $c_2 = \{\texttt{Name}, \texttt{Illness}\}$:

R	SSN	Name	Illness	HurtBy	Doctor
	?	Hellmann	Borderline	?	White

The Impact of A Priori Knowledge: Example (2)

Attacker's updated view on original instance r:

R	SSN	Name	Illness	HurtBy	Doctor
	?	Hellmann	Borderline	?	White
	?	Dooley	?	?	Warren
	?	McKinley	?	?	Warren
	?	McKinley	?	?	Warren

Suppose attacker knows a priori:
"All patients suffering from `Borderline` have hurt themselves."

As a sentence of $\mathscr{L}$:
$(\forall X_S)(\forall X_N)(\forall X_H)(\forall X_D)\,[\,R(X_S, X_N, \mathtt{BLine}, X_H, X_D) \Rightarrow (X_N \equiv X_H)\,]$

Attacker's updated view on r violates $c_3 = \{\mathtt{Name}, \mathtt{HurtBy}\}$:

R	SSN	Name	Illness	HurtBy	Doctor
	?	Hellmann	Borderline	Hellmann	White

About Inference-Proofness and A Priori Knowledge

Inference-Proofness: From attacker's point of view

- For each potential secret $\Psi_i(\boldsymbol{v_i}) \in \mathrm{ex}(potsec(\mathcal{C}))$
- Existence of alternative instance r' over $\langle R|A_R|SC_R\rangle$ possible
  - r' is indistinguishable from original instance r
  - r' does not satisfy $\Psi_i(\boldsymbol{v_i})$

About a priori knowledge *prior*

- Contains sentences over predicate symbols R and $\equiv$
- Attacker knows: Original instance r satisfies *prior*
- Consequently: Each r' also needs to satisfy *prior*

Towards Inference-Proofness of Alternative Instance

Create inference-proof alternative instance r' w.r.t.

- **Single** potential secret $\Psi_i(\boldsymbol{v_i})$ with $\boldsymbol{v_i} = (v_{i_1}, \ldots, v_{i_\ell})$
  - Attacker knows from f_1: $\pi_{(A_{F_1} \setminus \mathcal{E})}(r)$
  - Choose $m \in \{i_1, \ldots, i_\ell\}$ s.t. $a_m \notin (A_{F_1} \setminus \mathcal{E})$ (i.e. $a_m \in \bar{A}_{F_2}$)
  - Make sure: Column a_m of r' does not contain $v_m \in \boldsymbol{v_i}$
- Syntactically restricted sentence $\Gamma \in$ *prior* over R and $\equiv$
  - Attacker knows: Γ is satisfied by r
  - Adopt all columns $\{a_1, \ldots, a_n\} \setminus \{a_m\}$ of r to construct r'
  - Ensure that Γ does **not** require
    - Constant v_m to be in m-th column
    - Equality between column m and other column

A Priori Knowledge and Multiple Potential Secrets

Consider example set $\mathcal{C}$ within $\langle R|A_R|SC_R\rangle$

R	SSN	Name	Illness	HurtBy	Doctor
c_1	×				
c_2		×	×		
c_3		×		×	
c_4			×	×	

- Columns `Name` and `Doctor` known from f_1
  $\rightarrow$ Do **not** modify to preserve indistinguishability
- For each $\Psi_i(\boldsymbol{v_i})$: To be able to construct r' protecting $\Psi_i(\boldsymbol{v_i})$ at least one column of c_i must be modifiable
- Each $\Gamma \in$ *prior* must comply with all modifiable columns
  - In each $(\neg)R(\ldots)$ of Γ: No constants in modifiable columns
  - No equalities expressed by variables between modifiable and non-modifiable columns

technische universität dortmund

Definition of A Priori Knowledge

Each $\Gamma \in$ *prior* is built s.t.

- Γ has form $(\forall \boldsymbol{x})(\exists \boldsymbol{y})[\bigvee_{j=1,\ldots,p} \neg R(t_{j,1},\ldots,t_{j,n}) \vee A_{p+1}]$
  - A_{p+1} is either $(t_{p+1,1} \equiv t_{p+1,2})$ or $\bigwedge_{j=p+1,\ldots,q} R(t_{j,1},\ldots,t_{j,n})$
  - Each $t_{j,i}$ is a variable or a constant symbol
- Γ is range-restricted: Each $X \in \boldsymbol{x}$ occurs in a $\neg R(\ldots)$
- Γ is not DB-tautologic: No $Y \in \boldsymbol{y}$ occurs in a $\neg R(\ldots)$

technische universität dortmund

Definition of A Priori Knowledge

Moreover: *prior* must comply with "modifiable columns"

There exists a subset $M \subseteq \{h+1, \ldots, n\}$ s.t.

(1) $M \cap \{i_1, \ldots, i_\ell\} \neq \emptyset$ for each $c_i = (a_{i_1}, \ldots, a_{i_\ell}) \in \mathcal{C}$

(2) For each $\Gamma \in \mathit{prior}$ exists a partioning $\mathcal{X}_1^\Gamma \mathbin{\dot\cup} \mathcal{X}_2^\Gamma = \mathit{Var}$ s.t.

(i) For each atom $R(t_1, \ldots, t_n)$ of Γ

- For $j \notin M$: term t_j is either a (quantified) variable of $\mathcal{X}_1^\Gamma$ or a constant symbol of *Dom*
- For $j \in M$: term t_j is a (quantified) variable of $\mathcal{X}_2^\Gamma$

(ii) For each atom $(X_i \equiv X_j)$ of Γ:
Either $X_i, X_j \in \mathcal{X}_1^\Gamma$ or $X_i, X_j \in \mathcal{X}_2^\Gamma$

(iii) For each atom $(X_i \equiv v)$ of Γ with $v \in \mathit{Dom}$:
Variable X_i is in $\mathcal{X}_1^\Gamma$

Coarse Sketch of Proof

To be shown:
for all $\Psi(\boldsymbol{v}) \in \mathrm{ex}(\mathit{potsec}(\mathcal{C}))$: $db_{f_1} \cup db_R \cup \mathit{prior} \not\models_{DB} \Psi(\boldsymbol{v})$

Steps of proof:

1. Choose $\tilde{\Psi}(\boldsymbol{v}) \in \mathrm{ex}(\mathit{potsec}(\mathcal{C}))$ arbitrarily
2. Construct a DB-Interpretation $\mathcal{I}_{r'}$ with

$$\mathcal{I}_{r'} \models_M \begin{cases} db_{f_1} \\ db_R \\ \mathit{prior} \end{cases} \qquad \text{(Indistinguishability)}$$

$$\mathcal{I}_{r'} \not\models_M \tilde{\Psi}(\boldsymbol{v}) \qquad \text{(Non-violation of potential secret)}$$

Creation of Appropriate Fragmentation

Alternative Fragmentation of Example Instance

R	SSN	Name	Illness	HurtBy	Doctor
	1234	Hellmann	Borderline	Hellmann	White
	2345	Dooley	Laceration	McKinley	Warren
	3456	McKinley	Laceration	Dooley	Warren
	3456	McKinley	Concussion	Dooley	Warren

F_1	tid	SSN	Illness	HurtBy	Doctor
	1	e_S^1	Borderline	e_H^1	White
	2	e_S^2	Laceration	e_H^2	Warren
	3	e_S^3	Laceration	e_H^3	Warren
	4	e_S^4	Concussion	e_H^4	Warren

F_2	tid	SSN	HurtBy	Name
	1	κ_S^1	κ_H^1	Hellmann
	2	κ_S^2	κ_H^2	Dooley
	3	κ_S^3	κ_H^3	McKinley
	4	κ_S^4	κ_H^4	McKinley

is confidential w.r.t.

$$\mathcal{C} = \{ \quad c_1 = \{\mathtt{SSN}\}, \quad c_3 = \{\mathtt{Name}, \mathtt{HurtBy}\}, \quad c_2 = \{\mathtt{Name}, \mathtt{Illness}\}, \quad c_4 = \{\mathtt{Illness}, \mathtt{HurtBy}\} \quad \}$$

A Priori Knowledge under Alternative Fragmentation

Attacker's view on r based on f_1:

R	SSN	Name	Illness	HurtBy	Doctor
	?	?	Borderline	?	White
	?	?	Laceration	?	Warren
	?	?	Laceration	?	Warren
	?	?	Concussion	?	Warren

Suppose attacker knows a priori:

1. $(\forall X_S)(\forall X_N)(\forall X_I)(\forall X_H)\ [\,R(X_S, X_N, X_I, X_H, \texttt{White}) \Rightarrow (X_I \equiv \texttt{BLine})\,]$
2. $(\forall X_S)(\forall X_N)(\forall X_H)(\forall X_D)\,[\,R(X_S, X_N, \texttt{BLine}, X_H, X_D) \Rightarrow (X_N \equiv X_H)\quad]$

A Priori Knowledge is harmless (though premises satisfied)

1. Association `Doctor` $\leftrightarrow$ `Illness` already known from f_1
2. For neither X_N nor X_H a constant is known

About the Creation of Appropriate Fragmentations

As seen in example: Given $\langle R|A_R|SC_R\rangle$, $\mathcal{C}$ and *prior*
Some fragmentations achieve inference-proofness, others do not

Task: Create inference-proof fragmentation for given setting

- Can be modelled as Binary Integer Linear Program
- Optimization Goal: Minimize number of "encrypted attributes"
- Solver outputs feasible solution iff
  Inference-proof fragmentation exists

Conclusion and Future Work

Conclusion and Future Work

What has been achieved?

- Existing approach to confidentiality by fragmentation is
  - Modelled logic-orientedly within CIE-framework
  - Extended by attacker's a priori knowledge
- Within modelling: Formal proof of inference-proofness
- Algorithm for computing inference-proof fragmentations

What might be done in future?

- Extending feasible a priori knowledge
  $\rightarrow$ Sufficient & necessary condition
- Analyzing other approaches to confidentiality by fragmentation

That's all...

Thank you for your attention!