technische universität dortmund

Untersuchungen zum Inferenzschutz von fragmentierten Speicherungen von Datenbankinstanzen

Marcel Preuß

Lehrstuhl Informatik 6

Technische Universität Dortmund

23. November 2010

Ein kurzer Überblick

Vertraulichkeit durch Fragmentierung

- Grundlagen zur Fragmentierung
- Fragmentierung und partielle lokale Verwaltung

Inferenzsicherheit der fragmentierten Speicherung

- Vorgehensweise für Untersuchungen zur Inferenzsicherheit
- Wahl des logischen Systems
- Logik-orientierte Modellierung der (fragm.) Relationeninstanz
- Inferenzsicherheit unter Funktionalen Abhängigkeiten

Vertraulichkeit durch Fragmentierung

Wieso Fragmentierung?

Oft: Nur Assoziationen zwischen Informations-Aspekten sensibel

Beispiel: Welcher Patient wegen welcher Krankheit behandelt?

- Behandelte Krankheiten $\rightsquigarrow$ komplett unspannend
- Behandelte Patienten $\rightsquigarrow$ bedingt sensibel
- Assoziation: Patient und Krankheit $\rightarrow$ hochgradig sensibel

Grundsätzliche Idee der Fragmentierung

Ziel: Sensible Assoziationen durch Fragmentierung aufbrechen

- Betrachte Relationeninstanz r zu Schema $\langle R | A_R | \quad \rangle$
- Attributmenge A_R in Teilmengen zerlegen
  - Fragmentierung $\mathcal{F} = \{\langle F_1 | A_{F_1} | SC_{F_1} \rangle, \ldots, \langle F_n | A_{F_n} | SC_{F_n} \rangle\}$
  - Fragment $\langle F_i | A_{F_i} | SC_{F_i} \rangle$ ist Relationenschema mit $A_{F_i} \subseteq A_R$
- Bilde Projektionen von r gemäß der Fragmente
  $\rightarrow$ sogenannte Fragment-Instanzen $f_1, f_2, \ldots, f_n$

Beispiel-Instanz

Patient	<u>SSN</u>	Name	Geburtstag	Plz	Krankheit	Arzt
	12345	Hellmann	03.01.1981	94142	Bluthochdruck	White
	98765	Dooley	07.10.1953	94141	Fettleibigkeit	Warren
	24689	McKinley	12.02.1952	94139	Bluthochdruck	White
	13579	Ripley	03.01.1981	94139	Fettleibigkeit	Warren

Abbildung: Instanz *patient* zu Schema *Patient*

Was fällt auf?

- Attribut `SSN` ist Primärschlüssel
- sensible Assoziationen enthalten

Fragmentierung der Beispiel-Instanz

F_1	Name
	Hellmann
	Dooley
	McKinley
	Ripley

F_2	Geburtstag	Plz
	03.01.1981	94142
	07.10.1953	94141
	12.02.1952	94139
	03.01.1981	94139

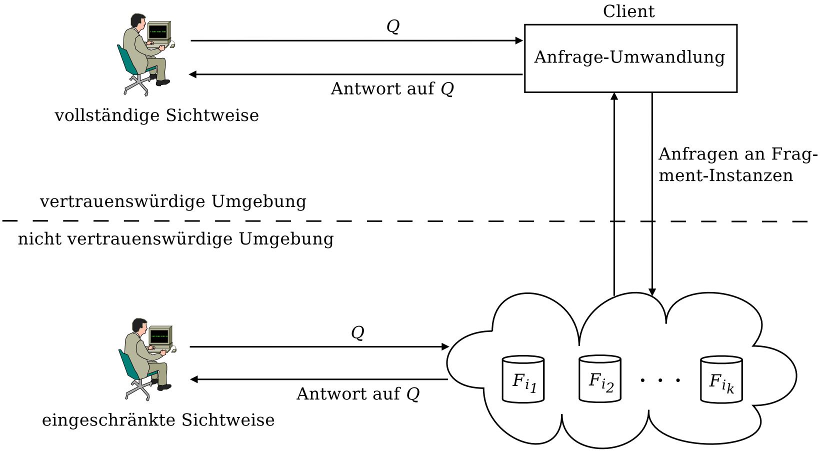
F_3	Krankheit	Arzt
	Bluthochdruck	White
	Fettleibigkeit	Warren

Abbildung: Mögliche Fragment-Instanzen f_1, f_2 und f_3 von *patient*

Was fällt auf?

- Primärschlüssel `SSN` in keinem Fragment enthalten
- sensible Assoziationen aufgebrochen

Konzept: Vertraulichkeit durch Fragmentierung



Ein konkreter Fragmentierungs-Ansatz

Konkreter Ansatz: Fragmentierung und partielle lokale Verwaltung

Getroffene Annahmen:

- Externe Server nicht vertrauenswürdig, aber ehrlich
- Client ist vollkommen vertrauenswürdig
- Client bietet (begrenzten) lokalen Speicherplatz
- lokale Verwaltung von Daten teurer als externe Verwaltung
  $\rightarrow$ bevorzugt externen Speicher nutzen

Aufbau der Fragmentierung

Zu den Annahmen passende Fragmentierung?

- Zwei Fragmente $\langle F_o | A_{F_o} | SC_{F_o} \rangle$ und $\langle F_s | A_{F_s} | SC_{F_s} \rangle$
  - Instanzen zu F_o ausschließlich lokal verwalten
  - Instanzen zu F_s können extern gespeichert werden
- Jedes Attribut in mindestens einem Fragment enthalten
- Rekonstruktion der ursprünglichen Instanz über Tupel-IDs
  $\rightarrow$ als Primärschlüssel in SC_{F_o} und SC_{F_s} vereinbart

Beispiel-Fragmentierung

F_o	tid	SSN	Name	Geburtstag
	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	Plz	Krankheit	Arzt
	1	94142	Bluthochdruck	White
	2	94141	Fettleibigkeit	Warren
	3	94139	Bluthochdruck	White
	4	94139	Fettleibigkeit	Warren

Abbildung: Fragmentierung von *patient* bei partieller lokaler Verwaltung

Definition von Vertraulichkeits-Anforderungen

Wie Vertraulichkeits-Anforderungen **formal** definieren?

Idee: Vertraulichkeits-Constraints

- $\langle R|A_R| \quad \rangle$ sei gegebenes Relationenschema
- Vertraulichkeits-Constraint c ist Teilmenge $c \subseteq A_R$
- Vertraulichkeits-Constraint c heißt
  - einfach, falls $|c| = 1$
  - assoziierend, falls $|c| > 1$

Semantik von Vertraulichkeits-Constraints

Gegeben sei:

- Relationenschema $\langle R|A_R| \quad \rangle$
- Menge $\mathcal{C}$ von Vertraulichkeits-Constraints,
- Fragmentierung $\mathcal{F} = \{\langle F_o|A_{F_o}|SC_{F_o}\rangle, \langle F_s|A_{F_s}|SC_{F_s}\rangle\}$ zu R

$\mathcal{F}$ heißt korrekt bezüglich $\mathcal{C}$ $\quad\Leftrightarrow\quad$ für jedes $c \in \mathcal{C}$ gilt: $c \not\subseteq A_{F_s}$

Beispiel: Vertraulichkeits-Constraints

$$\begin{aligned}
c_0 &= \{\texttt{SSN}\}\\
c_1 &= \{\texttt{Name}, \texttt{Geburtstag}\}\\
c_2 &= \{\texttt{Name}, \texttt{Plz}\}\\
c_3 &= \{\texttt{Name}, \texttt{Krankheit}\}\\
c_4 &= \{\texttt{Name}, \texttt{Arzt}\}\\
c_5 &= \{\texttt{Geburtstag}, \texttt{Plz}, \texttt{Krankheit}\}\\
c_6 &= \{\texttt{Geburtstag}, \texttt{Plz}, \texttt{Arzt}\}
\end{aligned}$$

Abbildung: Menge $\mathcal{C}$ von Vertraulichkeits-Constraints für *Patient*

Beispiel-Fragmentierung, korrekt bezüglich $\mathcal{C}$

F_o	tid	SSN	Name	Geburtstag
	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	Plz	Krankheit	Arzt
	1	94142	Bluthochdruck	White
	2	94141	Fettleibigkeit	Warren
	3	94139	Bluthochdruck	White
	4	94139	Fettleibigkeit	Warren

Abbildung: Fragmentierung von *patient* bei partieller lokaler Verwaltung

Inferenzsicherheit der fragmentierten Speicherung

Vorgehensweise für Untersuchungen zur Inferenzsicherheit

- CQE ist als inferenzsicher nachgewiesen
- Formalisiere Fragmentierung im CQE-Framework
  - Logisches System wählen
  - Logik-orientierte Modellierung der (fragmentierten) Relationeninstanz
  - Logik-orientierte Modellierung der Vertraulichkeits-Politik
- Formuliere formale Beweise zur Inferenzsicherheit
  - Annahmen zu potentiellen Angreifern
  - Annahmen zum Vorwissen von potentiellen Angreifern

Logisches System des CQE-Framework: Syntax

Syntax des verwendeten logischen Systems ($\rightarrow$ Sprache $\mathscr{L}$)

- Prädikatenlogik 1. Ordnung mit Gleichheit
  - Prädikatensymbol R mit Stelligkeit $|A_R|$
  - ausgezeichnetes (binäres) Prädikatensymbol $=$
  - unendlich große, feste Domäne *Dom*
    $\rightarrow$ Konstantenzeichen des DB-Schemas
  - unendlich große Variablen-Menge $Var := \{X_1, X_2, \dots\}$
- in einer Formel $R(t_1, \dots, t_{|A_R|})$ ist t_i Konstante oder Variable
- nur geschlossene Formeln konstruierbar
  $\rightarrow$ ausschließlich quantifizierte Variablen $(\forall, \exists)$

Logisches System des CQE-Framework: Semantik

Eine Interpretation $\mathcal{I}$ für $\mathscr{L}$ ist eine DB-Interpretation $\Leftrightarrow$

- Universum $\mathcal{U}$ = Domäne *Dom*
- für alle $v \in Dom$ gilt: $\mathcal{I}(v) = v$
- für R gilt: $\mathcal{I}(R) \subset \underbrace{\mathcal{U} \times \ldots \times \mathcal{U}}_{|A_R| \text{ mal}}$ ist endlich groß
- für $=$ gilt: $\mathcal{I}(=) = \{(v, v) \mid v \in \mathcal{U}\}$

$\mathcal{I}$ kann als vollständige Datenbankinstanz aufgefasst werden!

Sichtweise auf Fragment-Instanz f_s

- Annahme: ursprüngliche DB zwar vollständig
- Aber: A_R nicht komplett in $\langle F_s | A_{F_s} | SC_{F_s} \rangle$ enthalten
- Sichtweise: Fragment-Instanz f_s als Teil von r
  $\rightarrow$ Werte zu A_{F_s} bekannt, Werte zu $A_R \setminus A_{F_s}$ unbekannt

Noch einmal graphisch verdeutlicht

R	a_1	a_2	a_3
	$\mu_1[a_1]$	$\mu_1[a_2]$	$\mu_1[a_3]$
	$\mu_2[a_1]$	$\mu_2[a_2]$	$\mu_2[a_3]$
	$\mu_3[a_1]$	$\mu_3[a_2]$	$\mu_3[a_3]$

(a) Ursprüngliche Instanz r

F	a_1	a_2
	$\mu_1[a_1]$	$\mu_1[a_2]$
	$\mu_2[a_1]$	$\mu_2[a_2]$
	$\mu_3[a_1]$	$\mu_3[a_2]$

(b) Fragment-Instanz f

$R(F)$	a_1	a_2	a_3
	$\mu_1[a_1]$	$\mu_1[a_2]$	?
	$\mu_2[a_1]$	$\mu_2[a_2]$	?
	$\mu_3[a_1]$	$\mu_3[a_2]$	?

(c) f im Kontext von r

Modellieren der ursprünglichen Instanz: Positives Wissen

Gegeben sei:

- ursprüngliche Relationeninstanz r
- Relationenschema $\langle R|A_R| \quad \rangle$ mit $A_R = \{a_1, \ldots, a_{|A_R|}\}$

Positives Wissen aus r in $\mathscr{L}$:

$$db_r^+ := \left\{ R(v_1, \ldots, v_{|A_R|}) \mid \exists \mu \in r : \bigwedge_{i=1}^{|A_R|} \mu[a_i] = v_i \right\}$$

Positives Wissen: Beispielhafte Modellierung

Person	**Name**	**Geburtstag**	**Plz**
	Hellmann	03.01.1981	94142
	McKinley	12.02.1952	94139
	Ripley	03.01.1981	94139

$$db^{+}_{person} = \{ \quad \begin{array}{l} \mathit{Person}(\text{ Hellmann}, \quad 03.01.1981, \quad 94142\ ), \\ \mathit{Person}(\text{ McKinley}, \quad 12.02.1952, \quad 94139\ ), \\ \mathit{Person}(\text{ Ripley}, \quad 03.01.1981, \quad 94139\ ) \end{array} \quad \}$$

Vollständigkeit der Datenbankinstanz

Problem: Ein potentieller Angreifer weiß aber noch mehr

- Ursprüngliche Instanz r als vollständig angenommen
- jede nicht aufgezählte Wertekombination gilt auch nicht in r
  $\rightarrow$ Wissen der Form $\neg R(v_1, \ldots, v_{|A_R|})$
- Problem: Domäne unendlich groß $\rightarrow$ nicht explizit aufzählbar
- Rettende Idee: Completeness-Sentence

Modellieren der ursprünglichen Instanz: Negatives Wissen

Erstmal beispielhaft:

$$
\begin{array}{lllll}
(\forall X_N)(\forall X_G)(\forall X_P)\ [& & & & \\
(X_N = \text{Hellmann} & \wedge & X_G = 03.01.1981 & \wedge & X_P = 94142) \ \vee \\
(X_N = \text{McKinley} & \wedge & X_G = 12.02.1952 & \wedge & X_P = 94139) \ \vee \\
(X_N = \text{Ripley} & \wedge & X_G = 03.01.1981 & \wedge & X_P = 94139) \ \vee \\
\neg Person(X_N, X_G, X_P) & & & &]
\end{array}
$$

Und jetzt formal als Completeness-Sentence in db_r^-:

$$(\forall X_1) \ldots (\forall X_{|A_R|}) \left[\bigvee_{\mu \in r} \left(\bigwedge_{j=1}^{|A_R|} (X_j = \mu[a_j]) \right) \vee \neg R\left(X_1, \ldots, X_{|A_R|}\right) \right]$$

Logik-orientierte Modellierung von r

Damit: Prädikatenlogische Modellierung von r klar

$$db_r := db_r^+ \cup db_r^-$$

Aber: Potentieller Angreifer kennt nur Fragment f_s
$\rightarrow$ Modelliere f_s in Framework für r

Neue (kleinere) Beispiel-Instanz

Person	**Name**	**Geburtstag**	**Plz**
	Hellmann	03.01.1981	94142
	McKinley	12.02.1952	94139
	Ripley	03.01.1981	94139

Abbildung: Instanz *person* zum Relationen-Schema *Person*

Menge $\mathcal{C}$ von Vertraulichkeits-Constraints für Schema *Person*

- $c_0 = \{Name, Geburtstag\}$
- $c_1 = \{Name, Plz\}$

Fragmentierung der neuen Beispiel-Instanz

F_o	**tid**	**Name**
	1	Hellmann
	2	McKinley
	3	Ripley

F_s	**tid**	**Geburtstag**	**Plz**
	1	03.01.1981	94142
	2	12.02.1952	94139
	3	03.01.1981	94139

Abbildung: Mögliche Fragment-Instanzen f_o und f_s zu *person*

Positives Wissen aus f_s: Grundsätzliche Ideen

Potentieller Angreifer kennt $\langle F_s | A_{F_s} | SC_{F_s} \rangle$ und $\langle R | A_R | \quad \rangle$
$\rightarrow$ er weiß: Werte zu $A_R \setminus A_{F_s}$ werden ihm vorenthalten

Rückschlüsse von f_s auf r über positives Wissen

- f_s entspricht (bis auf `tid`) Projektion von r auf A_{F_s}
- Also: für $\mu \in f_s$ existiert (genau) ein $\nu \in r$ mit $\nu\lceil A_{F_s} = \mu\lceil A_R$

Modellieren der Fragment-Instanz f_s: Positives Wissen

Angreifer kennt Werte für $A_{F_s} = \{a_{i_1}, \ldots, a_{i_k}\} \subseteq \{a_1, \ldots, a_{|A_R|}\}$

Bilde Index-Mengen für (un)bekannte Attribut-Werte

- $\mathrm{Ind}^+_{F_s} = \{i_1, \ldots, i_k\}$
- $\mathrm{Ind}^-_{F_s} = \{1, \ldots, |A_R|\} \setminus \{i_1, \ldots, i_k\} = \{i_{k+1}, \ldots, i_{|A_R|}\}$

Konstruktion von $db^+_{f_s}$:

- $\forall \mu \in f_s$: füge $(\exists X_{i_{k+1}}) \ldots (\exists X_{i_{|A_R|}})\, R(t_1, \ldots, t_{|A_R|})$ hinzu
- Dabei gilt für $j \in \{1, \ldots, |A_R|\}$:
  - falls $j \in \mathrm{Ind}^+_{F_s}$: $t_j = \mu[a_j]$
  - falls $j \in \mathrm{Ind}^-_{F_s}$: $t_j = X_j$

Positives Wissen: Beispielhafte Modellierung

F_s	tid	Geburtstag	Plz
	1	03.01.1981	94142
	2	12.02.1952	94139
	3	03.01.1981	94139

$$
\begin{aligned}
db^+_{f_s} = \{ \quad & (\exists X_N)\ Person(\ X_N, \quad 03.01.1981, \quad 94142\ ), \\
& (\exists X_N)\ Person(\ X_N, \quad 12.02.1952, \quad 94139\ ), \\
& (\exists X_N)\ Person(\ X_N, \quad 03.01.1981, \quad 94139\ ) \quad \}
\end{aligned}
$$

Altbekanntes Problem: negatives Wissen

Problem: User weiß noch mehr

- Datenbankinstanz als vollständig angenommen
- für $\mu \notin f_s$: kein Tupel ν mit $\nu\lceil A_{F_s} = \mu$ in r
  $\rightarrow$ Wissen der Form $(\forall X_{i_{k+1}}) \ldots (\forall X_{i_{|A_R|}})\ \neg R(t_1, \ldots, t_{|A_R|})$
- altes Problem: unendliche große Domäne *Dom*
- alter Bekannter: Completeness-Sentence

Modellieren der Fragment-Instanz f_s: Negatives Wissen

Zur Erinnerung:

- $\mathrm{Ind}^{+}_{F_s} = \{i_1, \dots, i_k\}$
- $\mathrm{Ind}^{-}_{F_s} = \{1, \dots, |A_R|\} \setminus \{i_1, \dots, i_k\} = \{i_{k+1}, \dots, i_{|A_R|}\}$

Modifizierter Completeness-Sentence für $db^{-}_{f_s}$:

$$(\forall X_1)\dots(\forall X_{|A_R|})\left[\bigvee_{\mu \in f_s}\left(\bigwedge_{j \in \mathrm{Ind}^{+}_{F_s}} (X_j = \mu[a_j])\right) \vee \neg R\,(X_1, \dots, X_{|A_R|})\right]$$

Negatives Wissen: Beispielhafte Modellierung

F_s	tid	Geburtstag	Plz
	1	03.01.1981	94142
	2	12.02.1952	94139
	3	03.01.1981	94139

$$\begin{aligned}
&(\forall X_N)(\forall X_G)(\forall X_P)\ [\\
&(X_G = 03.01.1981 \ \wedge\ X_P = 94142)\ \vee\\
&(X_G = 12.02.1952 \ \wedge\ X_P = 94139)\ \vee\\
&(X_G = 03.01.1981 \ \wedge\ X_P = 94139)\ \vee\\
&\neg Person(X_N, X_G, X_P)\ ]
\end{aligned}$$

Wahl des Typs der Vertraulichkeits-Politik

Vertraulichkeits-Politik als Potential Secrets modellieren

- Annahme zum Ziel von Vertraulichkeits-Constraints: **existierende** Werte bzw. Assoziationen schützen
- modellieren als Potential Secrets $pot\_sec(\mathcal{C})$
  - falls Ψ_i wahr in DB: Benutzer darf dies *nicht* erfahren
  - sonst: Benutzer darf dies erfahren
- Annahme: Potentielle Angreifer kennen $\mathcal{C}$ bzw. $pot\_sec(\mathcal{C})$

Überlegungen zur Modellierung der Vertraulichkeits-Politik

Grundsätzliche Idee:

- Betrachte Constraint $c_i = \{a_{i_1}, \ldots, a_{i_k}\}$
- alle Kombinationen zugehöriger Werte schützen
- alle Kombinationen als Potential Secrets definieren
- leider unendlich viele wegen $|Dom| = \infty$
- benutze **freie** Variablen $X_{i_1}, \ldots, X_{i_k}$ für $a_{i_1}, \ldots, a_{i_k}$
  $\rightarrow$ endlich große Repräsentation

Logik-orientierte Modellierung der Vertraulichkeits-Politik

Betrachte Vertraulichkeits-Constraint $c_i = \{a_{i_1}, \ldots, a_{i_k}\}$ aus $\mathcal{C}$

- $\mathrm{Ind}_{c_i}^{+} = \{i_1, \ldots, i_k\}$
- $\mathrm{Ind}_{c_i}^{-} = \{1, \ldots, |A_R|\} \setminus \{i_1, \ldots, i_k\} = \{i_{k+1}, \ldots, i_{|A_R|}\}$

Konstruktion von $pot\_sec(\mathcal{C})$

- $\forall c_i \in \mathcal{C}$: konstruiere

$$\Psi_i(\boldsymbol{X_i}) = (\exists X_{i_{k+1}}) \ldots (\exists X_{i_{|A_R|}})\, R(X_1, \ldots, X_{|A_R|})$$

- Dabei gilt für $j \in \{1, \ldots, |A_R|\}$:
  - falls $j \in \mathrm{Ind}_{c_i}^{+}$: X_j ist freie Variable
  - falls $j \in \mathrm{Ind}_{c_i}^{-}$: X_j ist gebundene Variable
- $\boldsymbol{X_i} = (X_{i_1}, \ldots, X_{i_k})$ ist Vektor der freien Variablen

Expansion der Vertraulichkeits-Politik

Gegeben: $\Psi_i(\boldsymbol{X_i})$ mit $\boldsymbol{X_i} = (X_{i_1}, \ldots, X_{i_k})$

Konstruiere $\text{ex}(\Psi_i(\boldsymbol{X_i}))$:

- bilde jede Konstantenkombination $\boldsymbol{v_i} = (v_{i_1}, \ldots, v_{i_k})$
- bilde Formel $\Psi_i(\boldsymbol{v_i}) \in \text{ex}(\Psi_i(\boldsymbol{X_i}))$

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

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

Vernachlässigtes Vorwissen

Jetzt bekannt: Modellierung der Sichtweise eines potentiellen Angreifers auf eine fragmentierte DB-Instanz

Aber: Potentielle Angreifer können auch Vorwissen haben
$\rightarrow$ Vorwissen + DB-Wissen $\rightsquigarrow$ Inferenzen?

In CQE: Vorwissen explizit durch $prior \subset \mathscr{L}$ modelliert

Bisherige Resultate

Resultate zur Inferenzsicherheit laut Diplomarbeit:

- inferenzsicher, wenn $\mathit{prior} = \emptyset$
- nicht inferenzsicher, wenn Vorwissen „beliebig“,
  aber: $\forall\, \Psi(\boldsymbol{v}) \in \mathrm{ex}(\mathit{pot\_sec}(\mathcal{C})) :\ \mathit{prior} \not\models_{DB} \Psi(\boldsymbol{v})$
- inferenzsicher unter eingeschränkten Funktionalen Abh. (FDs)

Ziel: „Spielraum“ für Vorwissen erweitern!

Neues Resultat: Inferenzsicherheit unter **allgemeinen** FDs

Prädikatenlogische Modellierung von FDs

Betrachte: Relationenschema $\langle R|A_R|SC_R\rangle$ mit $SC_R = \Sigma$

FD $\in \Sigma$: $\underbrace{\{a_{e_1}, \ldots, a_{e_\ell}\}}_{\subseteq A_R} \rightarrow \underbrace{\{a_e\}}_{\in A_R}$

In $prior_\Sigma$:

$$(\forall X_1)\ldots(\forall X_{|A_R|})(\forall Y_1)\ldots(\forall Y_{|A_R|})[\ \begin{array}{l} (R(X_1,\ldots,X_{|A_R|}) \wedge \\ R(Y_1,\ldots,Y_{|A_R|}) \wedge \\ X_{e_1} = Y_{e_1} \wedge \ldots \wedge X_{e_\ell} = Y_{e_\ell}) \\ \Rightarrow X_e = Y_e \end{array}\ ]$$

Inferenzsicherheit unter FDs

Gegeben sei:

- ursprüngliche Instanz r zu
- Schema $\langle R|A_R|SC_R\rangle$ mit $A_R = \{a_1, \ldots, a_{|A_R|}\}$
- Menge $\mathcal{C}$ von Vertraulichkeits-Constraints zu $\langle R|A_R|SC_R\rangle$
- $\mathcal{F} = \{\langle F_o|A_{F_o}|SC_{F_o}\rangle, \langle F_s|A_{F_s}|SC_{F_s}\rangle\}$, korrekt bzgl. $\mathcal{C}$
- f_o und f_s Fragment-Instanzen zu $\mathcal{F}$ bzgl. r
- $db_{f_s} := db^+_{f_s} \cup db^-_{f_s}$ zu f_s als Sichtweise eines Angreifers
- $pot\_sec(\mathcal{C})$ als Vertraulichkeits-Politik
- $prior_\Sigma$ zu $SC_R = \Sigma$ als Vorwissen eines Angreifers

Zu zeigen: $\forall\, \Psi(\boldsymbol{v}) \in \mathrm{ex}(pot\_sec(\mathcal{C})) :\ prior_\Sigma \cup db_{f_s} \not\models_{DB} \Psi(\boldsymbol{v})$

Beweisskizze

Zu zeigen: $\forall\, \Psi(\boldsymbol{v}) \in \mathrm{ex}(pot\_sec(\mathcal{C}))$: $prior_{\Sigma} \cup db_{f_s} \not\models_{DB} \Psi(\boldsymbol{v})$

Ablauf des Beweises:

1. Wähle $\tilde{\Psi}(\boldsymbol{v}) \in \mathrm{ex}(pot\_sec(\mathcal{C}))$ beliebig
2. Zeige: es existiert ein $\mathcal{I}^*$ mit
   - $\mathcal{I}^* \models_M prior_{\Sigma}$
   - $\mathcal{I}^* \models_M db_{f_s}$
   - $\mathcal{I}^* \not\models_M \tilde{\Psi}(\boldsymbol{v})$

Im Folgenden: für $j \in \{1, \ldots, |A_R|\}$ gilt $j \in \mathrm{Ind}^{+}_{F_s} \iff$
Term t_j ist Konstante in beliebigem $\Phi \in db^{+}_{f_s}$

Semantischer Korrektheitsbeweis (1)

Über Formeln aus $db^{+}_{f_s}$:

- betrachte $\tilde{\Psi}(\boldsymbol{v}) \in \text{ex}(pot\_sec(\mathcal{C}))$ mit $\boldsymbol{v} = (v_{i_1}, \ldots, v_{i_p})$
- es gilt: $\tilde{\Psi}(\boldsymbol{X}) \in pot\_sec(\mathcal{C})$ mit $\boldsymbol{X} = (X_{i_1}, \ldots, X_{i_p})$
- weiter: $c = \{a_{i_1}, \ldots, a_{i_p}\} \in \mathcal{C}$
- Fragmentierung $\mathcal{F}$ ist korrekt bzgl. $\mathcal{C}$
  - also: $c = \{a_{i_1}, \ldots, a_{i_p}\} \not\subseteq A_{F_s}$
  - demnach: $\exists m \in \{i_1, \ldots, i_p\} : a_m \notin A_{F_s}$
- folglich: für alle Formeln aus $db^{+}_{f_s}$ gilt

$$\ldots(\exists X_m)\ldots \; R(t_1, \ldots, t_{|A_R|}) \quad \text{mit} \quad t_m := X_m$$

- insbesondere: $m \notin \text{Ind}^{+}_{F_s}$

Semantischer Korrektheitsbeweis (2)

Ziel des Beweises: DB-Interpretation $\mathcal{I}^*$

Hilfsweise: Konstruiere DB-Interpretation $\mathcal{I}_r$ mit $\mathcal{I}_r \models_M db_r$

- für alle $R(u_1, \ldots, u_{|A_R|}) \in db_r^+$: $(u_1, \ldots, u_{|A_R|}) \in \mathcal{I}_r(R)$
- keine weiteren Tupel in $\mathcal{I}_r(R)$

Eigenschaften von $\mathcal{I}_r$:

- offensichtlich: $\mathcal{I}_r \models_M db_r^+$
- offensichtlich: $\mathcal{I}_r \models_M db_r^-$
- $\mathcal{I}_r \models_M prior_\Sigma$, weil r alle FDs aus Σ beachtet

Semantischer Korrektheitsbeweis (3)

Idee: konstruiere gewünschtes $\mathcal{I}^*$ aus $\mathcal{I}_r$

- für alle $(u_1, \ldots, u_m, \ldots, u_{|A_R|}) \in \mathcal{I}_r(R)$:
  füge $(u_1, \ldots, \varphi_m(u_m), \ldots, u_{|A_R|})$ in $\mathcal{I}^*(R)$ ein
- keine weiteren Tupel in $\mathcal{I}^*(R)$

$\varphi_m : \mathcal{U}_m \to \mathcal{U} \setminus \{v_m\}$ ist **injektive** Funktion, wobei

- für alle $(u_1, \ldots, u_m, \ldots, u_{|A_R|}) \in \mathcal{I}_r(R)$: $u_m \in \mathcal{U}_m$
- $\mathcal{U}$ ist (unendlich großes) Universum von $\mathcal{I}^*$
- v_m ist aus $\boldsymbol{v} = (v_{i_1}, \ldots, v_{i_p})$

φ_m kann stets konstruiert werden, weil $||(\mathcal{U} \setminus \{v_m\})|| > ||\mathcal{U}_m||$

Semantischer Korrektheitsbeweis (4)

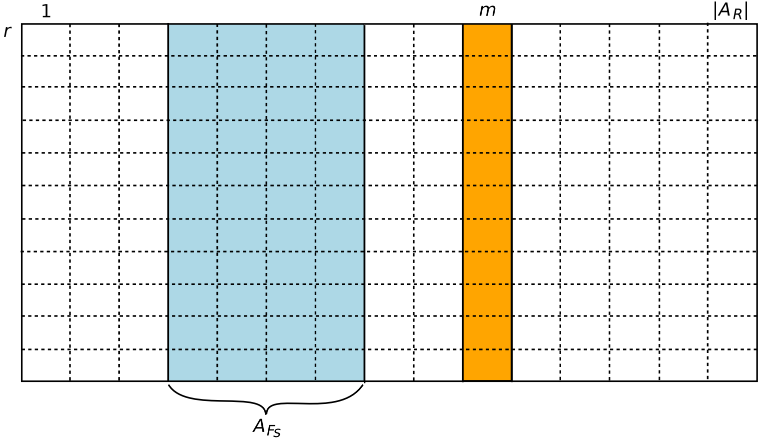
Ziel: für $\tilde{\Psi}(\boldsymbol{v}) \in \mathrm{ex}(pot\_sec(\mathcal{C}))$ gilt $prior_{\Sigma} \cup db_{f_s} \not\models_{DB} \tilde{\Psi}(\boldsymbol{v})$

Zwischenstand: $\mathcal{I}^*$ konstruiert

Noch zu zeigen: $\mathcal{I}^*$ erfüllt die gewünschten Eigenschaften

- $\mathcal{I}^* \models_M db_{f_s}$
- $\mathcal{I}^* \models_M prior_{\Sigma}$
- $\mathcal{I}^* \not\models_M \tilde{\Psi}(\boldsymbol{v})$

Semantischer Korrektheitsbeweis (5)



Es gilt: $\mathcal{I}^* \models_M db_{f_s}$ wegen $\mathcal{I}_r \models_M db_r$ und $m \notin \mathrm{Ind}^+_{F_s}$

Semantischer Korrektheitsbeweis (6)

Noch zu zeigen: $\mathcal{I}^* \models_M \mathit{prior}_\Sigma$

Formel aus prior_Σ:

$$(\forall X_1)\ldots(\forall X_{|A_R|})(\forall Y_1)\ldots(\forall Y_{|A_R|})[\ (R(X_1,\ldots,X_{|A_R|}) \wedge R(Y_1,\ldots,Y_{|A_R|}) \wedge X_{e_1} = Y_{e_1} \wedge \ldots \wedge X_{e_\ell} = Y_{e_\ell}) \Rightarrow X_e = Y_e\ ]$$

Auffällig: nur spaltenweise Gleichheiten entscheidend!
(spaltenweise, weil „getypte" Formeln in prior_Σ)

Diese Gleichheiten aus $\mathcal{I}_r$ müssen in $\mathcal{I}^*$ **genau** erhalten bleiben
$\rightarrow$ auch keine „neuen" Gleichheiten

Semantischer Korrektheitsbeweis (7)

Betrachte aus $\mathcal{I}_r(R)$:
$(u_1, \ldots, u_m, \ldots, u_{|A_R|})$ und $(w_1, \ldots, w_m, \ldots, w_{|A_R|})$

Dafür in $\mathcal{I}^*(R)$:
$(u_1, \ldots, \varphi_m(u_m), \ldots, u_{|A_R|})$ und $(w_1, \ldots, \varphi_m(w_m), \ldots, w_{|A_R|})$

Bleiben **alle** Gleichheiten **spaltenweise** erhalten?

- für $j \in \{1, \ldots, |A_R|\} \setminus \{m\}$: offensichtlich ja
- für Index m:
  - für $u_m = w_m$ gilt: $\varphi_m(u_m) = \varphi_m(w_m)$ (weil Funktion)
  - für $u_m \neq w_m$ gilt: $\varphi_m(u_m) \neq \varphi_m(w_m)$ (wegen Injektivität)

Semantischer Korrektheitsbeweis (8)

Es bleibt noch zu zeigen: $\mathcal{I}^* \not\models_M \tilde{\Psi}(\boldsymbol{v}) \quad$ mit $\quad \boldsymbol{v} = (v_{i_1}, \ldots, v_{i_p})$

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

- es existiert ein $(u_1, \ldots, u_m, \ldots, u_{|A_R|}) \in \mathcal{I}^*(R)$, in dem
- für **alle** $j \in \{i_1, \ldots, i_p\} : u_j = v_j$

Das gilt aber nicht, weil

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

q.e.d.

Das war es...

Vielen Dank für eure Aufmerksamkeit!