WEAKLY INDEPENDENT DATABASE SCHEMES

Ke Wang

Department of Information Systems and Computer Science

National University of Singapore, Singapore, 0511

Weining Zhang, Siu-Cheung Chau

Department of Mathematics and Computer Sciences

The University of Lethbridge

Lethbridge, Alberta, Canada, T1K 3M4

Abstract

Independent database schemes are best known to have constraints enforced globally by enforcing them locally in individual relations, and therefore are highly desirable in an updates-busy and distributed environment. However, independence excludes all database schemes with interrelational functional dependencies and therefore is too strong a condition. In this paper, a natural generalization of independent schemes along this line, called *weakly independent schemes*, is defined. A database scheme is weakly independent with respect to a set of embedded functional dependencies if the constraints can be enforced globally by enforcing local as well as interrelational functional dependencies. (a) A characterization of the weak independence and an exponential time test (in the number of functional dependencies) are presented. (b) Two polynomial time conditions are also presented, one is necessary and the other is sufficient for a database scheme to be *weakly independent*. (c) It is shown that a certain acyclicity of the database scheme implies weak independence. Extension to the case of embedded functional dependencies plus the join dependency $\bowtie \mathbf{R}$ is also considered.

Streams: database designs, database integrity, database theory and algorithms

Key Words: functional dependencies, independent schemes, uniqueness condition, weak instances

1 Introduction

The *universal instance assumption* or the *weak instance theory* was first studied by Honeyman [14] and independently by Vassiliou [26] as a means of defining the satisfaction of functional dependencies by a state, i.e., a collection of relations. In this assumption, a state is considered to be satisfying or *consistent* with respect to a set of functional dependencies if no contradictory information is inferred from the state by applying certain rules associated with the semantics of the functional dependencies [2, 20]. This process of inferences, called *chase* in the literature, is made through the use of nulls, which naturally imposes an expensive cost on consistency checking because the chase is a tuple-at-a-time computation. Such a computation is particularly intolerable when states are large and updates of states are frequent. The question "can we use the universal instance assumption without using nulls?" came in Sagiv's mind [21] and investigation into this issue [13, 16, 22] has come up with the notion of *independent database schemes*. [1] A database scheme is independent with respect to a set of functional dependencies if every state in which each relation satisfies its local functional dependencies is consistent. For an independent scheme, inferences on nulls are not needed for checking consistency and consequently constraints can be enforced efficiently.

Unfortunately, all database schemes having interrelational functional dependencies are excluded from being independent, where a functional dependency is interrelational if it is embedded in two relation schemes. There are realistic database designs in which such dependencies are needed. With the presence of interrelational functional dependencies, an "explicit violation" occurs when two tuples, either from a single relation as for a local functional dependency or from two different relations as for an interrelational functional dependency, agree on all attributes of the left-hand side and disagree on some attribute of the right-hand side. Since the tuples to be examined for checking explicit violation of functional dependencies can be extracted by projections, explicit violations can be as easily checked as violations of local satisfaction. If this checking is sufficient for checking the global consistency, the expensive chase can be avoided. These observations lead us to identify a more general class of database schemes, called *weakly independent schemes* in this paper. Informally, a database scheme is weakly independent if the absence of explicit violations of functional dependencies implies consistency. Independent schemes form the subclass of weakly independent schemes that have no interrelational functional dependencies. When location of data, i.e., whether within a single relation or in different relations, makes no "significant" differences on the access efficiency, independence is too strong a condition; weakly independent schemes are just what are needed.

---

[1] However, some primitive ideas of independence go back to as early as late 70's [5, 8].

The following results are obtained. (1) A characterization of weak independence and an exponential time test (in the number of functional dependencies) are presented. Since database schemes are small (compared with states) and scheme design is a "once for all" matter while updates of states could be frequent, our strategy of sacrificing design time to gain efficiency on constraint enforcement and therefore on updates should make sense. (2) Two polynomial time conditions are also presented, one is necessary and the other is sufficient for weak independence. These conditions are essentially modifications of the characterization. (3) We will show that a certain acyclicity of the database scheme implies weak independence. Consequently, when no interrelational functional dependencies are present, such an acyclicity implies independence, establishing a relationship between two well-known database properties. Therefore, while the acyclicity has been known to be desirable for query answering and optimization [7, 11, 28], our results show it is also desirable for constraint enforcement and updates. All results are obtained in the general case that functional dependencies are embedded in the relation schemes. Extension to the case of embedded functional dependencies plus the join dependency $\bowtie \mathbf{R}$ is also considered. Interestingly, by considering the special case where no interrelational functional dependencies are present, which is necessary for independence, all our characterizations and conditions are reduced immediately to the *uniqueness condition* of independent schemes [23, 24].

When only key dependencies are considered, independent schemes are necessarily in BCNF [22]. This is no longer true for weakly independent schemes. Example 1 in Section 3 provides a weakly independent scheme that is in 3NF but not in BCNF. In view of the fact that BCNF may be too strong, weak independence serves a more general design goal in this respect. Weakly independent schemes form a subclass of *constant-time-maintainable (ctm)* schemes studied by Graham and Wang [12, 27], i.e., schemes for which validation of updates of states can be checked in time independent of the state size. While both classes have exponential time tests when only embedded functional dependencies appear, the test for weakly independent schemes is more "manageable" in the following sense. The cover of functional dependencies used in the weak independence test has a size polynomially related to the size of the original cover (see the definition of the cover G in Section 3). On the other hand, the test for ctm schemes requires, in general, an embedded cover of exponential size, in which dependencies embedded in each relation scheme are equivalent to all implied dependencies embedded in that scheme [12, 27]. Moreover, in the ctm scheme test, tableaux are examined for a certain property which is more complicated than computing derivations. This will make weakly independent schemes more easily adoptable to existing systems than ctm schemes.

2   Preliminaries

We now describe the notation and definitions required for the rest of this paper. The standard notation can be found in [18, 25].

2.1   Relations, Schemes, and States

There are a finite number of *attributes* or *columns*, each being associated with a set of constants called the *domain*. We denote attributes using the capital letters from the beginning of the alphabet (e.g., $A, B, C, A_1, B_1, C_1, \ldots$). Following relational database theory notation, we represent the union of two sets X and Y of attributes as XY, the intersection as $X \cap Y$, and the difference as $X - Y$. A *relation scheme R* is a non-empty set of attributes. A *tuple* over a relation scheme R is a mapping from attributes of R to constants in domains of attributes of R, such that each attribute $A \in R$ is mapped to a constant in the domain of A. A *relation* over a relation scheme R is a finite set of tuples over R. We use $\bowtie$ and $\prod_X()$ for *join* and *projection* operations of relations, respectively [10].

A *(database) scheme*, denoted $(\mathbf{R}, \Sigma)$, consists of a collection of relation schemes $\mathbf{R} = \{R_1, \ldots, R_m\}$ and a finite set of dependencies Σ over $\cup\mathbf{R}$ defined below, where $\cup\mathbf{R}$ refers to the abbreviation of the union $R_1 \cup \ldots \cup R_m$. Sometimes, we call $\mathbf{R}$ a database scheme without mentioning functional dependencies Σ. A *(database) state* over $\mathbf{R}$, usually denoted ρ, is an assignment of relations to relation schemes of $\mathbf{R}$, with $\rho(R_i)$ denoting the relation assigned to R_i by ρ. It is also useful to treat a state as a set of tuples over elements of $\mathbf{R}$ and, therefore the statements like "$\rho' \subseteq \rho$" and "a tuple in ρ" should make sense.

2.2   Functional Dependencies

A *functional dependency (fd)* [1, 10] over a set W of attributes is a statement of the form $X \rightarrow Y$, where X and Y are sets of attributes from W, and they are called the *left-hand side (lhs)* and *right-hand side (rhs)* of the fd, respectively. Semantically, a relation $\rho(R_i)$ over R_i *satisfies* an fd $X \rightarrow Y$ over R_i if whenever there exist two tuples u and v in $\rho(R_i)$ such that $u[X] = v[X]$, then $u[Y] = v[Y]$. An fd $X \rightarrow Y$ is *trivial* if $X \supseteq Y$; otherwise, it is *non-trivial*. We shall consider only non-trivial fd's as trivial fd's do not act as constraints. A set of fd's F *(logically) implies* a set of fd's F', denoted $F \models F'$, if whenever a relation r satisfies F, then r also satisfies F'. $F \not\models F'$ means that $F \models F'$ does not hold. If $F \models G$ and $G \models F$, F is said to be *equivalent* to (or to be a *cover* of) G, and we denote this by $F \equiv G$. F^+ is the set of all fd's implied by F. Let X be a set of attributes. X_F^+ is the set of attributes A such that $F \models X \rightarrow A$. X_F^+ can be computed in linear time in the size of the description of F [4]. An fd $X \rightarrow Y$ is *embedded* in a relation scheme R_i if

$R_i \supseteq XY$. F/R_i denotes the fd's of F that are embedded in R_i. F is embedded in $\mathbf{R}$ if every fd in F is embedded in some element of $\mathbf{R}$. Finally, for any two fd's, $X_1 \to Y_1$ and $X_2 \to Y_2$, we say $X_1 \to Y_1$ is *embedded* in $X_2 \to Y_2$ if $X_1Y_1 \subseteq X_2Y_2$.

Let F be a set of fd's. Given two sets U and V of attributes, a *derivation* of U from V with respect to (wrt) F is a finite sequence of fd's $X_1 \to Y_1 \in F, \ldots, X_n \to Y_n \in F$, $n \geq 0$, such that $U \subseteq VY_1 \ldots Y_n$ and, for all $1 \leq j \leq n$, $X_j \subseteq YY_1 \ldots Y_{j-1}$. A derivation of U from V wrt F is *minimal* if no fd in the derivation can be removed such that the resulting sequence is still a derivation of U from V wrt F.

2.3 Chase and Consistency

Given a state ρ over $\mathbf{R}$, we define a tableau $aug_U(\rho)$ over $\cup\mathbf{R}$ and call it the *tableau for state ρ*: For each relation scheme $R_i \in \mathbf{R}$ and for each tuple $t \in \rho(R_i)$, there is a row s in $aug_U(\rho)$ corresponding to it. The row is the result of padding t as follows: $s[R_i] = t$; and $s[A]$, for all $A \in \cup\mathbf{R} - R_i$, is a *null* that represents some unknown value and appears no else where in $aug_U(\rho)$. Let $(\mathbf{R}, \Sigma)$ be a database scheme and ρ a state over $\mathbf{R}$. ρ is a *consistent state* of $(\mathbf{R}, \Sigma)$, or is *consistent* wrt Σ, if there exists a relation I over $\cup\mathbf{R}$ satisfying Σ such that $\prod_{R_i}(I) \supseteq \rho(R_i)$ for each R_i in $\mathbf{R}$ [14, 26, 28]. Such a relation I is called a *weak instance* of ρ. Constraint enforcement is to enforce consistency of database states.

We can test whether a database state ρ is consistent wrt a set F of fd's by applying the *chase* process [2, 20] to $aug_U(\rho)$, the state tableau for ρ. In general, the chase process modifies a tableau T by applying the following fd-rules to T as far as possible, until either a conflict occurs, or no rule can further modify the tableau.

> *fd-rule:* for each $X \to A$, there is an fd-rule corresponding to it. Suppose tableau T has rows t_1, t_2 that agree on all X-columns. Let v_1, v_2 be the values in the column A of t_1, t_2, respectively. Furthermore, assume $v_1 \neq v_2$. Applying the fd-rule corresponding to $X \to A$ to rows t_1, t_2 of T yields a new tableau T'. T' is the same as T except v_1, v_2 are renamed as follows: If one of v_1 and v_2 is a null, then rename this null by the other. If both are distinct constants, then t_1 and t_2 *conflict* for $X \to A$. For an fd $X \to Y$, where Y is a set of attributes, the effect of applying the fd-rule for $X \to Y$ is the effect of applying fd-rules for $X \to A$ for all $A \in Y$.

A state ρ is consistent wrt a set F of fd's if and only if no conflict occurs in applying the chase process to the tableau $aug_U(\rho)$ using fd-rules associated with fd's in F. Enforcing constraints by direct use of the chase process is expensive and time consuming.

5

2.4 Independence

Graham and Yannakakis [13], and Sagiv [21, 22] defined and studied the notion of independent schemes. Let $(\mathbf{R}, F)$ be a database scheme, where $F = F_1 \cup \ldots \cup F_m$ and each F_i is a set of fd's embedded in $R_i \in \mathbf{R} = \{R_1, \ldots, R_m\}$. It was shown that the following two conditions are equivalent [13]: (a) every state ρ in which $\rho(R_i)$ satisfies F_i for every $R_i \in \mathbf{R}$, called a *locally satisfying* state, is consistent wrt F; (b) every state ρ in which $\rho(R_i)$ satisfies F^+/R_i for every $R_i \in \mathbf{R}$ is consistent wrt F. A database scheme $\mathbf{R}$ is *independent* wrt F if the scheme satisfies either condition. Therefore, for independent schemes, only local fd's need be enforced and this can be done much more efficiently than direct enforcement of consistency via the chase process. Note that $(\mathbf{R}, F)$ is not independent if a non-trivial fd in F^+ is embedded in two relation schemes, as a counterexample can always be constructed [13, 22, 23]. Polynomial time characterizations of independent schemes have been given in [13, 16, 23, 24].

3 Weakly Independent Schemes

In this section, we define the notion of weak independence and discuss its efficient constraint enforcement. The following examples show that independence is too strong a condition in a centralized environment.

Example 1 The design algorithm of Biskup *et al* [6] produces the scheme $\mathbf{R} = \{R_1(CAZ), R_2(ZC)\}$ and $F = F_1 \cup F_2$, where $F_1 = \{CA \to Z, Z \to C\}$, $F_2 = \{Z \to C\}$ (i.e., "City, Address, Zip"). Clearly, $\mathbf{R}$ is not independent wrt F because $Z \to C$ is embedded in both relations [13, 22]. On the other hand, for any given state ρ, if $\rho(R_i)$ satisfies F_i, $i = 1, 2$, and the set $\prod_{R_1 \cap R_2}(\rho(R_1)) \cup \prod_{R_1 \cap R_2}(\rho(R_2))$ satisfies the "interrelational" fd $Z \to C$, then ρ is consistent wrt F. Therefore, when a state is updated, we need only check satisfaction of the local and "interrelational" fd's involved.□

The above example is simple in illustrating the idea, but is artificial in that it has two relation schemes such that one is contained in the other. Now we consider a more "realistic" example that is not independent but shares the same property as the above example.

Example 2 An immigration database may consist of the following three relations:

R_1: Immi-file(Name,Age,Sex,File#,Mail-addr,Details)

R_2: Reg-file(Name,Age,Sex,File#,Apply-date)

R_3: Visa-file(Name,Age,Sex,Mail-addr,Passport#,Visa-type)

- The relation Immi-file contains detailed information of the applicants who are applying for permanent residence or citizenship. Some of the information is confidential (as for Details) and therefore only the officer (but not the receptionist) may have access.

- Relation Reg-file is mainly used by the front desk receptionist to arrange interviews. Information in this file is less confidential than Immi-file.

- The relation Visa-file is used for issuing visas by the visa office. A person who just has a short visit rather than applies for immigrant status will have a record in this file, but not in Immi-file.

Also, attributes are distinguished by their names, not by where they appear (i.e., the *universal relation scheme assumption* [19]). In particular, the File# in both relations denotes the same application file number for immigrant status. There are the following fd's: $G = G_1 \cup G_2 \cup G_3$, where

$$G_1: \text{File\#} \rightarrow \text{Name,Age,Sex,Mail-addr,Details,}$$
$$\text{Name,Age,Sex} \rightarrow \text{File\#,Mail-addr,Details;}$$
$$G_2: \text{File\#} \rightarrow \text{Name,Age,Sex,Apply-date,}$$
$$\text{Name,Age,Sex} \rightarrow \text{File\#,Apply-date;}$$
$$G_3: \text{Name,Age,Sex} \rightarrow \text{Mail-addr,Passport\#,Visa-type.}$$

For the same reason as in Example 1, **R** is clearly not independent wrt G.

We now consider interrelational fd's that are to be enforced on the common attributes of every pair of relation schemes.

$$\text{Name,Age,Sex} \rightarrow \text{File\#} \quad \text{and} \quad \text{File\#} \rightarrow \text{Name,Age,Sex}$$

will be the interrelational fd's for pair R_1, R_2. The interrelational fd's for pair R_1, R_3 will be

$$\text{Name,Age,Sex} \rightarrow \text{Mail-addr.}$$

There are no interrelational fd's for pair R_2, R_3. The formal definition of interrelational fd's will be given shortly. We claim that the scheme has the following property: if a state ρ satisfies local as well as interrelational fd's, then ρ is consistent wrt G. A formal proof of this claim will not be given until we have defined the notion of weak independence and presented the test algorithm. Right now, however, we can make an essential and informal observation in order to give some intuitive idea behind it: in this scheme, a violation of the uniqueness condition of [23], that is, the existence of a derivation σ_1 of XA, for some $X \rightarrow Y \in G_j$ and some $A \in Y$ from R_i with $R_j \neq R_i$, implies the existence of a derivation σ_2 of A from X wrt either fd's strictly weaker than $X \rightarrow Y$ in the sense of [23] or interrelational fd's that are embedded in those fd's used in derivation σ_1. This essentially

says that an inconsistency of a state implies a violation of either local or interrelational fd's. This observation will stand on a more precise and formal ground when more notation is developed.□

Now we formalize the above idea. Let F be a set of fd's embedded in $\mathbf{R}$. For the rest of the paper, we will consider only the cover G of F in the following form. For each $R_i \in \mathbf{R} = \{R_1, \ldots, R_m\}$, we define

$$G_i = \{X \to X_F^+ \cap R_i - X \mid X \text{ is a lhs of } F,\ X \subseteq R_i, \text{ and } X_F^+ \cap R_i - X \neq \emptyset\}.$$

We assume that each fd in G_i is implicitly *tagged* by the relation scheme R_i and two fd's are considered to be the same if and only if their tags, lhs's, and rhs's are identical, respectively. Let $G = G_1 \cup \ldots \cup G_m$. $\{G_1, \ldots, G_m\}$ is a partition of G from our assumption on the identity of fd's. Clearly, the cover G can be obtained from F in polynomial time. It is easy to see the following facts:

Fact 0: $G \equiv F$. This is because for each fd $X \to Y \in F$ embedded in R_i, there is an fd $X \to W$ in G_i such that $W \supseteq Y$.

Fact 1: For any fd $X \to Y \in G_i$, XY is a set of attributes *locally closed* in R_i; that is, $(XY)_G^+ \cap R_i = XY$. Also, $X_{G_i}^+ = XY$.

Fact 2: $G_i \models F/R_i$. The argument is the same as Fact 0.

Fact 3: For any lhs X of F or G and an attribute $A \notin X$, if $X \to A \in F^+/R_i$, then there is exactly one fd $X \to Y \in G_i$ such that $A \in Y$. This is immediate from the definition of G_i. However, Fact 3 does not say that for every fd $X \to A \in F^+/R_i$, $G_i \models X \to A$. Example 3 below gives an instance where the implication $\models$ does not hold.

Now we define the *interrelational fd's* between R_i and R_j as

$$G(i,j) = \{X \to Y \cap R_i \cap R_j \mid X \to Y \in G_i \cup G_j,\ X \subseteq R_i \cap R_j,\ Y \cap R_i \cap R_j \neq \emptyset\}.$$

Note that $G(i,i) = G_i$, for all $R_i \in \mathbf{R}$; that is, local fd's in G_i are treated uniformly as interrelational fd's. For any independent scheme $(\mathbf{R}, F)$, $G(i,j) = \emptyset, i \neq j$ [13, 23, 24], where G is obtained from F as above. The reader should observe that by requiring each fd in G to be locally closed more interrelational fd's are captured in the cover G. From Fact 3 and definition of $G(i,j)$, we have

Fact 4: For any fd $W \to Z \in G(i,j)$, there exists exactly one fd $W \to Y \in G_i$ such that $Z \subseteq Y$; similarly, there exists exactly one fd $W \to X \in G_j$ such that $Z \subseteq X$. That is, each interrelational fd is also enforced within the corresponding relations. We will make use of this fact in later proofs without explicitly mentioning it.

Example 3 Consider database scheme $(\mathbf{R}, F)$, where

G_1	G_2	G_3	G_4	G_5
$CD \to B$	$AB \to C$	$D \to E$	$BE \to C$	$CH \to B$
$BC \to D$	$C \to A$	$D \to H$	$BC \to E$	$BC \to H$
	$BC \to A$	$CH \to D$	$CE \to B$	
		$CE \to H$		
		$D \to EH$		
		$CH \to DE$		
		$CE \to HD$		
		$C \to A$		

Figure 1: G_i for Example 3

$$\mathbf{R} = \{R_1(BCD), R_2(ABC), R_3(ACDEH), R_4(BCE), R_5(BCH)\}$$

and F contains

$$AB \to C$$
$$C \to A$$
$$BC \to D$$
$$CD \to B$$
$$D \to E$$
$$D \to H$$
$$BE \to C$$
$$CH \to D$$
$$CE \to H$$

G is computed in Figure 1. All $G(i,j)$, $i \neq j$, are empty except $G(2,3) = G(3,2) = \{C \to A\}$. □

Let R_i and R_j be relation schemes in $\mathbf{R}$, not necessarily distinct. We define a set of tuples

$$\rho(i,j) = \prod_{R_i \cap R_j}(\rho(R_i)) \cup \prod_{R_i \cap R_j}(\rho(R_j)).$$

In particular, if $R_i \cap R_j = \emptyset$ then $\rho(i,j) = \emptyset$; and if $i = j$ then $\rho(i,j) = \rho(R_i) = \rho(R_j)$. $\rho(i,j)$ is the set of tuples on which the interrelational fd's in $G(i,j)$ will be enforced. The following is the main definition.

Definition 1 A state ρ is said to be *embedded-satisfying* G if for every pair R_i, R_j in $\mathbf{R}$, not necessarily distinct, $\rho(i,j)$ satisfies $G(i,j)$. $\mathbf{R}$ is said to be *weakly independent* wrt G if every embedded-satisfying state is consistent wrt G. □

We make the following comments about these definitions.

- Given a weakly independent scheme $(\mathbf{R}, G)$ and a state ρ, consistency of ρ can be tested by checking the satisfaction of relation $\rho(i,j)$ wrt fd's $G(i,j)$, for every pair R_i, R_j in $\mathbf{R}$, where $\rho(i,j)$ can be extracted from ρ by two projections and one union. In the process of updates where the current state ρ is known to be consistent wrt G and a tuple t_i on some R_i is inserted into $\rho(R_i)$, consistency of $\rho \cup \{t_i\}$ can be tested more efficiently as follows: for every $R_j \in \mathbf{R}$ and every fd $X \to Y \in G(i,j)$, retrieve from relation $\rho(R_i)$ the tuples with $X-$value being $t_i[X]$ by one selection and check if the returned tuples have $t_i[Y]$ as its $Y-$value. Since all returned tuples must have the same $Y-$value (because ρ is consistent), the selection can be replaced by retrieval of a single tuple. Clearly, the number of such retrievals needed for checking this insertion is $\Sigma_{R_j \in \mathbf{R}} \mid G(i,j) \mid$, which is independent of the size of states, where $\mid G(i,j) \mid$ is the number of fd's in $G(i,j)$. Finally, note that deletion of tuples preserves consistency, so weakly independent schemes are desirable in a large and update-busy environment.

- The notion of embedded-satisfaction is dependent on the choices of covers of the original fd set F. In particular, let G be obtained from F and let G' be obtained from F' where $F \equiv F'$. Then the embedded-satisfaction wrt G may not imply the embedded-satisfaction wrt G'. Consider the following cover F' of F in Example 3:

$$AB \to C$$
$$C \to A$$
$$BC \to D$$
$$D \to E$$
$$D \to H$$
$$BE \to C$$
$$CH \to B$$
$$CE \to H$$

G' obtained from F' is computed in Figure 2. We can see that $G'_1 \not\models G_1$. Therefore, it is possible that some state embedded-satisfying G' does not embedded-satisfy G. In general, for two embedded covers F and F' where F contains all lhs's of F', the embedded-satisfaction wrt G obtained from F implies the embedded-satisfaction wrt G' obtained from F', and thus, the weak independence wrt G' implies the weak independence wrt G. Hence, the user can vary the strength or generality of weak independence by choosing different embedded covers of fd's. But there is a trade-off between the generality of weak independence and the efficiency of test and constraint enforcement because the latter depends on the size of G.

G_1'	G_2'	G_3'	G_4'	G_5'
$BC \to D$	$AB \to C$	$D \to E$	$BE \to C$	$CH \to B$
	$C \to A$	$D \to G$	$BC \to E$	$BC \to H$
	$BC \to A$	$CE \to H$	$CE \to B$	
		$D \to EH$		
		$CE \to HD$		
		$C \to A$		

Figure 2: G_i'

- The embedded-satisfaction excludes violations of fd's by tuples from a single relation or different relations. If the pair R_i, R_j in the above definition are required to be the same, or if $G(i,j) = \emptyset$ for all $i \neq j$, as required by independent schemes, then embedded-satisfaction reduces to local satisfaction and weakly independent schemes reduce to independent schemes. Thus, embedded-satisfaction implies local satisfaction and weak independence is a generalization of independence.

4   Characterization

In this section, we present a test for weakly independent database schemes. The test, however, is in exponential time in the number of fd's in G.

Let $X \to Y$ and $Q \to W$ be two fd's in G_i. The lhs X (or the fd $X \to Y$) is *weaker than* (*strictly weaker than, equivalent to*, respectively) the lhs Q (or the fd $Q \to W$) if $XY \subseteq QW$ ($XY \subset QW$, $XY = QW$, respectively). Note that these definitions are given for fd's having the same tag R_i ,i.e., contained in a single G_i, while embeddedness of fd's depends only on containment of attributes in fd's, not on tags. We shall use the following notation:

- $X_{w,i}^*$ — the closure of X wrt the fd's of G_i whose lhs's are strictly weaker than X.

- $G(X, R_i)$ — a set obtained by removing from G every fd of G_i (with tag R_i) having a lhs Y such that X is weaker than Y. Note that $G(X, R_i)$ contains all fd's in G_j, for $j \neq i$.

- $EQ(X, R_i)$ — the set of fd's in G_i whose lhs's are equivalent to X.

Given a database scheme $(\mathbf{R}, G)$, where G is in the form as in Section 3, our main result in

11

this section is: $\mathbf{R}$ is weakly independent wrt G if and only if every $R_i \in \mathbf{R}$ satisfies the following condition:

Weak Uniqueness Condition: Let $R_i \in \mathbf{R}$. R_i is said to satisfy the weak uniqueness condition wrt G if, for every fd $X \to Y \in G_j$ with $R_i \neq R_j$ and every attribute $A \in Y - X_{w,j}^*$, that a sequence $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ is a minimal derivation of XA from R_i wrt $G(X, R_j)$ implies $G(\sigma) \models X \to A$, where

$$G(\sigma) = (G(i,j)/XY) \ \cup \ (\cup_{k=1}^{l} G(q_k,j)/(X_k Y_k \cap XY)).$$

Intuitively, $G(\sigma)$ is the set of interrelational fd's that are "relevant" to derivation σ and are embedded in $X \to Y$ and in fd's used in σ. This condition essentially says that if there is a minimal derivation σ of XA from R_i wrt $G(X, R_j)$, then there is a derivation of A from X wrt $G(\sigma)$.

Example 4 Consider the following scheme:

$$\mathbf{R} = \{R_1(ABCD), R_2(ABE), R_3(BCH)\}$$
$$G = G_1 \cup G_2 \cup G_3, \text{ where}$$
$$G_1 = \{A \to BCD, B \to ACD\}$$
$$G_2 = \{A \to BE, B \to AE\}$$
$$G_3 = \{B \to HC\}$$

The interrelational fd's are as follows:

$$G(i,i) = G_i, \ 1 \leq i \leq 3$$
$$G(1,2) = G(2,1) = \{A \to B, B \to A\}$$
$$G(1,3) = G(3,1) = \{B \to C\}$$
$$G(2,3) = G(3,2) = \emptyset$$

Note that for every fd $X \to Y \in G_i, 1 \leq i \leq 3$, $X_{w,i}^* = X$, so $Y - X_{w,i}^* = Y$. Assume $i = 2, j = 1$. Let $X \to Y$ be $A \to BCD \in G_1$. $G(A, R_1) = \{A \to BE, B \to AE, B \to HC\}$. $\sigma_1 :<>$ is the only minimal derivation of AB from R_2 wrt $G(A, R_1)$; $\sigma_2 : B \to HC$ is the only minimal derivation of AC from R_2 wrt $G(A, R_1)$; there is no minimal derivation of AD from R_2 wrt $G(A, R_1)$. We have

$$G(\sigma_1) = G(1,2)/ABCD = \{A \to B, B \to A\};$$
$$G(\sigma_2) = G(1,2)/ABCD \ \cup \ G(3,1)/(BHC) \cap (ABCD)$$
$$= \{A \to B, B \to A, B \to C\}.$$

Clearly, $G(\sigma_1) \models A \to B$ and $G(\sigma_2) \models A \to C$. Similarly, we can show that the same property holds for every fd $X \to Y$ in G_j and every j such that $j \neq 2$. Therefore, R_2 satisfies the weak uniqueness condition wrt G. $\square$

Before we give the formal proof of the above characterization, some informal but intuitive arguments may help. Assume that the weak uniqueness condition holds for every R_i. As stated in the hypothesis of the condition, there are two ways to derive the attributes XA, i.e., the derivation σ and the empty derivation from R_j. A state ρ that contains a "two-derivations conflict" can be constructed from these two derivations in the natural way as in [22, 23, 24]. Since $G(\sigma) \models X \to A$, a conflict will occur in a special chase of ρ wrt $G(\sigma)$; in particular, this special chase applies an fd-rule for each fd in $G(q_k, j)$ (from $G(\sigma)$) to the tuples corresponding to $X_k \to Y_k$ and $X \to Y$, $1 \leq k \leq l$. Such a conflict corresponds to a violation of embedded-satisfaction of G because it occurs when an interrelational fd is applied to two tuples in the corresponding relations. Suppose now that $\mathbf{R}$ is not weakly independent wrt G. By Lemma 1 below, there exists a counterexample to the weak independence, i.e., an embedded-satisfying but inconsistent state, such that its inconsistency is caused by a "two-derivations conflict" as in ρ. Then the above comments imply that this counterexample is not embedded-satisfying G, a contradition. Therefore, $\mathbf{R}$ must be weakly independent wrt G. On the other hand, if the weak uniqueness condition is not satisfied for some R_i, then no conflict will occur in the above special chase (wrt $G(\sigma)$) of the above constructed state ρ, due to $G(\sigma) \not\models X \to A$. We can show that this implies that ρ is embedded-satisfying G and therefore is a counterexample to the weak independence (because ρ is inconsistent wrt G).

When $G(i,j) = \emptyset$ for all $i \neq j$, $G(\sigma)$ contains only fd's in G_j strictly weaker than $X \to Y$ and thus $G(\sigma) \not\models X \to A$. In this case, the weak uniqueness condition is reduced to the following uniqueness condition, as named by Sagiv, and we have the following proposition.

> *Uniqueness Condition [23, 24]:* Let $R_i \in \mathbf{R}$. R_i is said to satisfy the uniqueness condition wrt G if for every fd $X \to Y \in G_j$ with $R_j \neq R_i$ and every attribute $A \in Y - X_{w,j}^*$, $(R_i)_{G(X,R_j)}^+ \not\supseteq XA$. (That is, there exists no derivation of XA from R_i wrt $G(X, R_j)$)

Proposition 1 *(a) if $G(i,j) = \emptyset$ for all $i \neq j$, then $\mathbf{R}$ is weakly independent wrt G if and only if $\mathbf{R}$ is independent wrt G. (b) If every relation scheme contains exactly two attributes, then $\mathbf{R}$ is weakly independent wrt G if and only if $\mathbf{R}$ is independent wrt G.*□

Example 5 Consider a "Instructor, Course, Department" database scheme

$$\mathbf{R} = \{R_1(ID), R_2(IC), R_3(CD)\} \text{ and } G = \{I \to D, I \to C, C \to I\}.$$

$\mathbf{R}$ is not independent wrt G: the state $\rho = \{<i, d_1>, <i, c>, <c, d_2>\}$ is locally satisfying G, but is inconsistent wrt G. $\mathbf{R}$ is not weakly independent wrt G either by Proposition 1(a) or (b). In fact, state ρ is embedded-satisfying G.□

Now we prove that the weak uniqueness condition characterizes the weak independence. First, we borrow some notation from [16].

Let ρ be a state over $\mathbf{R}$. Consider a chase process of $aug_U(\rho)$ under G. If a tuple μ in $aug_U(\rho)$ is transformed into a tuple μ' by a number of applications of fd-rules for G, then μ is said to be *expanded* to μ', which is called an *expansion* of μ. We also say that μ' is an expansion of the tuple in ρ from which μ originates. An application of an fd-rule for $X \to Y \in G_i$ to μ and ν that agree on X columns is said to be *restricted* if either μ or ν is an expansion of a tuple in $\rho(R_i)$. If μ and ν conflict for a restricted application for $X \to Y \in G_i$, then the conflict is said to be *restricted*.

The following lemma is a modification of Lemma 1 in [16], which essentially says that if a scheme is not weakly independent, there exists a counterexample to weak independence in which inconsistency is caused by a "two-derivations conflict".

Lemma 1 *If $\mathbf{R}$ is not weakly independent wrt G, then there exists an embedded-satisfying state ρ such that a restricted conflict occurs by a number of restricted applications of fd-rules for G to $aug_U(\rho)$.*

The proof is a straightforward modification of the proof of Lemma 1 in [16]. Here we will consider embedded-satisfaction and weak independence rather than local satisfaction and independence. We first prove some claims. In the following claims, we assume that ρ is an embedded-satisfying state and define $aug_U(\rho)^*$ to be a tableau obtained by restrictedly applying fd-rules for G to $aug_U(\rho)$ until no nulls can be replaced with constants. Further assume that no restricted conflict occurs in $aug_U(\rho)^*$. Note that all nulls in $aug_U(\rho)^*$ are distinct. Let $X \to Y \in G_j$ and let μ and ν be tuples in $aug_U(\rho)^*$ that agree on columns X. Let μ have constants exactly over columns V (Then $X \subseteq V$). Define μ_j to be a tuple over R_j that agrees with μ on columns $V \cap (XY)$ and has distinct constants (that appear no where else in $aug_U(\rho)^*$) in all other columns.

Claim 1: $\rho' = \rho \cup \{\mu_j\}$ is embedded-satisfying G.

Proof of Claim 1: Assume that, for some $R_i \in \mathbf{R}$ and some tuple $\tau \in \rho(R_i)$, there exists an fd $Z \to W \in G(i,j)$ such that $\tau[Z] = \mu_j[Z]$. This implies that $Z \subseteq V \cap XY$ and thus τ agrees with μ on columns Z. Since no nulls can be replaced with constants by restricted applications of fd-rules for G to $aug_U(\rho)^*$, we must have $\mu[W] = \tau[W]$ (otherwise, from Fact 4 we can apply restrictedly an fd $Z \to Q$ in G_i with $W \subseteq Q$ to μ and τ, and replace some null in $\mu[W]$ with a constant in $\tau[W]$, a contradiction). Hence $W \subseteq V$ as all nulls in $aug_U(\rho)^*$ are distinct. Since $Z \subseteq XY$ and XY is locally closed, we have $W \subseteq XY$. So $W \subseteq V \cap (XY)$. From the definition of μ_j, $\mu_j[W] = \mu[W] = \tau[W]$. Claim 1 is proved.

Let $\rho' = \rho \cup \{\mu_j\}$. We can obtain $aug_U(\rho)^* \cup \{aug_U(\mu_j)\}$ by a number of restricted applications

of fd-rules for G to $aug_U(\rho')$, where $aug_U(\mu_j)$ is the tuple obtained by augmenting μ_j with unique nulls out to columns $\cup \mathbf{R}$. In the above context, we have the following claim.

Claim 2: (1) The fd-rule for $X \to Y$ can be restrictedly applied to μ and $aug_U(\mu_j)$. (2) If there is an attribute $A \in Y$ such that $\mu[A] \neq \nu[A]$ and $\nu[A]$ is a constant, then ν and $aug_U(\mu_j)$ restrictedly conflict for $X \to Y$; otherwise, fd-rule for $X \to Y$ can be restrictedly applied to ν and $aug_U(\mu_j)$.

Proof of Claim 2: Part (1) follows because $X \subseteq V \cap (XY)$ and (thus) $\mu[X] = aug_U(\mu_j)[X]$. Part (2) follows because $\nu[X] = \mu[X] = aug_U(\mu_j)[X]$ and $\mu_j[A]$ is either a new constant, if $A \notin V \cap XY$, or $\mu_j[A]$ is constant $\mu[A]$, if $A \in V \cap XY$. So Claim 2 follows.

Now we prove Lemma 1.

Proof of Lemma 1: Assume that $\mathbf{R}$ is not weakly independent wrt G. By the definition, there is an embedded-satisfying state ρ such that a conflict occurs in a chase process of $aug_U(\rho)$ under G. Without loss of generality, we consider a chase process that computes $aug_U(\rho)^*$ and then applies fd-rules for G to $aug_U(\rho)^*$ until a conflict occurs. Suppose that a conflict occurs by k applications of fd-rules for G to $aug_U(\rho)^*$. We prove the lemma by an induction on the number k.

Basis. $k = 0$. That is, a conflict occurs in $aug_U(\rho)^*$. If the conflict is restricted, then the lemma follows. So assume no restricted conflict occurs in $aug_U(\rho)^*$. Then there must be μ and ν in $aug_U(\rho)^*$ such that μ and ν agree on columns X but have distinct constants in a column in Y for some fd $X \to Y \in G_j$. Consider the state ρ' as defined in Claim 1, then the lemma follows from Claim 2.

Induction. $k \geq 1$. Let the first (non-restricted) application of fd-rules for G to $aug_U(\rho)^*$ be the one for some $X \to Y$ in G_j to μ and ν. Consider the state ρ' defined in Claim 1. When we compute $aug_U(\rho')^*$, by Claim 2, either ν and $aug_U(\mu_j)$ restrictedly conflict for $X \to Y$, or the first non-restricted application can be replaced with restricted applications of fd-rules for $X \to Y$ to μ and $aug_U(\mu_j)$ (and ν and $aug_U(\mu_j)$). Since $aug_U(\rho') = aug_U(\rho) \cup \{aug_U(\mu_j)\}$, for every tuple δ in $aug_U(\rho)^*$, there is an expansion of δ in $aug_U(\rho')^*$. Thus a conflict occurs by at most $k - 1$ applications of fd-rules for G to $aug_U(\rho')^*$. From the induction hypothesis, the lemma follows. □

Lemma 2 *Let $X \to Y \in G_j$ and $A \in Y - X_{w,j}^*$. Let $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ be a derivation of XA from R_i wrt $G(X, R_j)$, $R_i \neq R_j$. Let t be any tuple over $\cup \mathbf{R}$ and $\rho = \{v_i, u_1, \ldots, u_l, v_j\}$, where*

- *for $1 \leq k \leq l$, u_k is a tuple over R_{q_k} such that $u_k = t[X_k Y_k]$;*

- *v_i is a tuple over R_i such that $v_i = t[R_i]$;*

- v_j is a tuple over R_j such that $v_j[X] = t[X]$ and $v_j[A] \neq t[A]$.

If $G(\sigma) \models X \to A$, then ρ is not embedded-satisfying G, where

$$G(\sigma) = (G(i,j)/XY) \ \cup \ (\cup_{k=1}^{l} G(q_k, j)/(X_k Y_k \cap XY)).$$

Proof: Consider the following applications of fd's in $G(\sigma)$ to $aug_U(\rho)$: apply fd's in $G(i,j)/XY$ to $aug_U(v_i)$ and $aug_U(v_j)$, apply fd's in $G(q_k, j)/(X_k Y_k \cap XY)$ to $aug_U(u_k)$ and $aug_U(v_j)$. Since each fd applied is embedded in the relation schemes of both involved tuples, the effect of each application is either no change or a conflict on some column in the rhs of the fd. By $G(\sigma) \models X \to A$ and the construction of ρ, a conflict will eventually occur in the above applications. Then the first conflict occurs between some v_i or u_k and v_j, which is a violation of embedded-satisfaction.□

Lemma 3 *If every $R_i \in \mathbf{R}$ satisfies the weak uniqueness condition wrt G, then $\mathbf{R}$ is weakly independent wrt G.*

Proof: Suppose that $\mathbf{R}$ is not weakly independent wrt G. By Lemma 1, there is an embedded-satisfying state ρ in which a restricted conflict occurs in computing $aug_U(\rho)^*$. Assume that for the first time some tuples μ and ν conflict restrictedly for an fd $X \to Y \in G_j$; that is, one of μ and ν, say μ, is the expansion of a tuple in $aug_U(\rho(R_j))$ and, for some $A \in Y$, $\mu[X] = \nu[X]$ but $\mu[A]$ and $\nu[A]$ are distinct constants. We may assume that fd's are selected in the order of weakness of lhs's in computing $aug_U(\rho)^*$. Thus, at the time $X \to Y$ is selected for this restricted application to μ and ν, all fd's strictly weaker than $X \to Y$ have been selected for such applications. Hence, $A \in Y - X_{w,j}^*$.

Now consider the way ν was derived. Let ν be the expansion of a tuple $\tau_0 \in \rho(R_i)$ with $R_j \neq R_i$ and be derived by applying restrictedly fd-rules for $X_k \to Y_k \in G_{q_k}$, $1 \leq k \leq l$, to $aug_U(\tau_0)$ and tuples $\tau_k \in \rho(R_{q_k})$ in that order. Note that if $R_j = R_i$ or $l = 0$, then ρ is not embedded-satisfying G, because in that case $XA \subseteq R_i \cap R_j$ and $\{\mu[R_j \cap R_i], \tau_0[R_j \cap R_i]\} \subseteq \rho(i,j)$, which violates an fd $X \to W \in G(i,j)$ such that $A \in W$. Assume each of the above applications fills a constant in ν. Then the sequence $\sigma : X_1 \to Y_1, \ldots, X_l \to Y_l$ is a minimal derivation of XA from R_i wrt G. We claim that σ is also a minimal derivation of XA from R_i wrt $G(X, R_j)$; that is, for any X_k that is a lhs of G_j, $1 \leq k \leq l$, X is not weaker than X_k. Suppose X is weaker than some X_k. Then $XY \subseteq X_k Y_k$, $R_{q_k} = R_j$, and $\{\tau_k, \mu[R_j]\} \subseteq \rho(R_j)$. Also, since ν is derived by τ_k and $X_k \to Y_k$, we have $\tau_k[X] = \nu[X]$ and $\tau_k[Y] = \nu[Y]$. But $\mu[X] = \nu[X]$ and $\mu[Y] \neq \nu[Y]$. This implies that ρ is not embedded-satisfying G, a contradiction. The claim is proved.

From the weak uniqueness condition and the minimal derivation σ, we have $G(\sigma) \models X \to A$, where

$$G(\sigma) = (G(i,j)/XY) \cup (\cup_{k=1}^{l} G(q_k, j)/(X_k Y_k \cap XY)).$$

Since the restricted conflict between μ and ν is the first one occurring in computing $aug_U(\rho)^*$, no conflict occurs in the above derivation of constants of ν, so we have $\tau_k[X_kY_k] = \nu[X_kY_k]$, $1 \le k \le l$. Also, note that $\tau_0 = \nu[R_i]$, $\mu[X] = \nu[X]$ and $\mu[A] \ne \nu[A]$. Then from $G(\sigma) \models X \to A$ and Lemma 2, it follows that $\rho' = \{\tau_0, \tau_1, \ldots, \tau_l, \mu[R_j]\}$ is not embedded-satisfying G. Since $\rho' \subseteq \rho$, ρ is not embedded-satisfying G either. But this contradicts the fact established at the beginning of the proof.$\square$

Lemma 4 *Assume that state ρ contains only tuples that have constant 0's in locally closed sets of attributes, i.e., any set $X \subseteq R_i$ such that $X_G^+ \cap R_i = X$, and have distinct constants in other columns. Then ρ is embedded-satisfying G.*

Proof: Let u, v be two tuples in ρ over R_i, R_j, respectively. Further assume that there exists an fd $Q \to Z \in G(i,j)$ such that $u[Q] = v[Q]$. Then u, v have constants 0's over columns Q. Since $Q \to Z$ is embedded in both R_i and R_j, from the assumption of the lemma, u and v must have 0's over columns Z, and so $u[Z] = v[Z]$. $\square$

Lemma 5 *If $\mathbf{R}$ is weakly independent wrt G, then every $R_i \in \mathbf{R}$ satisfies the weak uniqueness condition wrt G.*

Proof: Suppose some R_i does not satisfy the weak uniqueness condition wrt G. Then for some fd $X \to Y \in G_j$ with $R_i \ne R_j$ and some attribute $A \in Y - X_{w,j}^*$, we have a minimal derivation $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ of XA from R_i wrt $G(X, R_j)$ such that $G(\sigma) \not\models X \to A$, where

$$G(\sigma) = (G(i,j)/XY) \cup (\cup_{k=1}^l G(q_k,j)/(X_kY_k \cap XY)).$$

Clearly, $XA \not\subseteq R_i$, otherwise, $G(i,j)/XY \models X \to A$ and $G(\sigma) \models X \to A$. Thus, $l \ge 1$. We now construct a state ρ as follows.

- for $1 \le k \le l$, $\rho(R_{q_k})$ contains a tuple u_k over R_{q_k} with constants 0's over columns X_kY_k and distinct constants over other columns;

- $\rho(R_i)$ contains a tuple v_i with constants 0's over all columns of R_i;

- $\rho(R_j)$ contains a tuple v_j over R_j with constants 0's over all columns $X_{G(\sigma)}^+$ and distinct constants over other columns. Note that $v_j[A]$ is a distinct constant because $A \notin X_{G(\sigma)}^+$.

Clearly, ρ is inconsistent wrt G because its construction enables the derivation σ to produce a contradiction on column A. We claim that ρ is embedded-satisfying G, and therefore $\mathbf{R}$ is not weakly independent wrt G. In fact, from Lemma 4, $\rho - \{v_j\}$ is embedded-satisfying G. So any possible violation of embedded-satisfaction must involve tuple v_j. There are the following two cases.

Case 1. There exists an fd $Z \rightarrow W \in G(i, j)$ such that $v_i[Z] = v_j[Z]$. Clearly, $Z \subseteq X^+_{G(\delta)} \subseteq XY$ from the construction of ρ, and thus $ZW \subseteq XY$ because $G \models Z \rightarrow W$ and XY is locally closed in R_j wrt G. Therefore, $Z \rightarrow W \in G(i,j)/XY \subseteq G(\sigma)$, and $ZW \subseteq X^+_{G(\sigma)}$ because $Z \subseteq X^+_{G(\sigma)}$. This shows $v_i[W] = v_j[W]$.

Case 2. There exists an fd $Z \rightarrow W \in G(q_k, j)$ for some $1 \leq k \leq l$ such that $u_{q_k}[Z] = v_j[Z]$. From the construction of ρ, $Z \subseteq (X_k Y_k) \cap X^+_{G(\sigma)} \subseteq (X_k Y_k) \cap (XY)$. Then we have $ZW \subseteq X_k Y_k$ and $ZW \subseteq XY$, because $X_k Y_k$ and XY are locally closed in R_{q_k} and R_j wrt G, respectively. Thus $Z \rightarrow W \in G(q_k, j)/(X_k Y_k \cap XY) \subseteq G(\sigma)$, and $ZW \subseteq X^+_{G(\sigma)}$ because $Z \subseteq X^+_{G(\sigma)}$. This and $ZW \subseteq X_k Y_k$ show that $u_{q_k}[W] = v_j[W]$.

In either cases, we have shown that ρ is embedded-satisfying G as required.□

Now we are ready to give the proof of the main theorem.

Theorem 1 (Main Theorem) **R** *is weakly independent wrt G if and only if every $R_i \in$* **R** *satisfies the weak uniqueness condition wrt G.*

Proof: This follows immediately from Lemmas 3 and 5.□

We have shown that the weak uniqueness condition characterizes the weak independence. Given an fd $X \rightarrow Y \in G_j$, an attribute $A \in Y - X^*_{w,j}$, and a derivation σ of XA from R_i wrt $G(X, R_j)$, $R_i \neq R_j$, it can be checked in polynomial time whether $G(\sigma) \models X \rightarrow A$. Therefore, we can test in non-deterministic polynomial time, thus in deterministic exponential time, whether a given scheme is *not* weakly independent. In the following, we present a deterministic exponential time test of weak independence that does not require to enumerate derivations.

Theorem 2 *(1) weak-unique(i)=*true *if and only if R_i satisfies the weak uniqueness condition wrt G. (2) The program test-weak-independence returns* true *if and only if* **R** *is weakly independent wrt G.*

Proof: Part (2) follows from Part (1) and Theorem 1. We show only Part (1).

(only if) Assume for some fd $X \rightarrow Y \in G_j$ with $R_j \neq R_i$, there exists a minimal derivation $\sigma : X_1 \rightarrow Y_1 \in G_{q_1}, \ldots, X_l \rightarrow Y_l \in G_{q_l}$ of XA from R_i wrt $G(X, R_j)$ for some $A \in Y - X^*_{w,j}$. Clearly, $X_k \rightarrow Y_k \notin G_i$, $1 \leq k \leq l$. Let $\alpha = \{X_1 \rightarrow Y_1, \ldots, X_l \rightarrow Y_l\}$ and let

$$\beta = G(\sigma) = (G(i,j)/XY) \cup \{\cup G(q_k, j)/(X_k Y_k \cap XY) \,|\, X_k \rightarrow Y_k \in \alpha\}.$$

Then $\alpha \subseteq G(X, R_j) - G_i$ and $(R_i)^+_\alpha \supseteq XA$. Since *weak-unique(i)=*true, $A \in X^+_\beta = X^+_{G(\sigma)}$. So $G(\sigma) \models X \rightarrow A$.

(if) For any fd $X \rightarrow Y \in G_j$ with $R_j \neq R_i$ and any non-empty subset α of $G(X, R_j) - G_i$ selected by the function *weak-unique(i)*, let

Input: A database scheme $\mathbf{R} = \{R_1, \ldots, R_m\}$ and a set of fd's $G = G_1 \cup \ldots \cup G_m$. Each G_i is a set of fd's embedded in R_i in the form as assumed in Section 3.

Output: **true** if the database scheme $\mathbf{R}$ is weakly independent wrt G; **false** otherwise.

```
program test-weak-independence;
var weak-independent: boolean;
function weak-unique(i: integer): boolean;
/* this function determines whether Ri satisfies the weak uniqueness condition wrt G */
begin
        weak-unique:=true;
       for each Rj ≠ Ri do
           for each fd X → Y ∈ Gj such that Y − X*w,j ≠ ∅ do
(*)             while not all non-empty subsets of G(X, Rj) − Gi are examined do
               begin
                       let α be an unexamined subset of G(X, Rj) − Gi;
                       let β = (G(i, j)/XY) ∪ {∪G(qk, j)/(XkYk ∩ XY) | Xk → Yk ∈ α is in Gqk};
                       if (A ∈ Y − X*w,j and (Ri)+α ⊇ XA) implies A ∉ X+β
(**)                   then weak-unique:=false;
           end;
   end;
   begin/* main program */
        weak-independent:=true;
       for each relation scheme Ri ∈ R do
       weak-independent:= weak-independent and weak-unique(i);
       write (weak-independent);
   end.
```

Figure 3: Test of Weak Independence

$$\beta = (G(i,j)/XY) \cup \{\cup G(q_k,j)/(X_kY_k \cap XY) \mid X_k \to Y_k \in \alpha \text{ is in } G_{q_k}\}.$$

Assume there exists some attribute $A \in Y - X^*_{w,j}$ such that $(R_i)^+_\alpha \supseteq XA$. There must be a minimal derivation σ of XA from R_i wrt α, and $\beta \models G(\sigma)$. From the weak uniqueness condition, $G(\sigma) \models X \to A$ and thus $A \in X^+_\beta$. Therefore, *weak-unique(i)*=**true**.□

The above program can run more efficiently by considering in the while loop only subsets of $G(X, R_j) - G_i$ that contain no two fd's $Q \to W$ and $U \to V$ such that $QW = UV$ because a derivation containing such fd's is obviously not minimal. Also, the test is in polynomial time for certain subclasses of database schemes and fd's. In principle, any database scheme with fd's that have only polynomially many derivations will have a polynomial time test because the complexity of the test is determined by the number of possible derivations.

Example 6 We now run the program on the scheme in Example 2. For simplicity and to save space, we will replace the attributes there by the following letters.

> File# — A
>
> Name, Age, Sex — B
>
> Mail-addr — C
>
> Details — D
>
> Apply-date — E
>
> Passport#,Visa-type — H

Note that the attribute groups (Name,Age,Sex) and (Passport#,Visa-type) are replaced by single letters B and H, respectively. No generality will be lost by such replacements because each group appears in the original scheme as a whole. The new scheme now is

> $\mathbf{R} = \{R_1(ABCD), R_2(ABE), R_3(BCH)\}$
>
> $G = G_1 \cup G_2 \cup G_3$, where
>
> $G_1 = \{A \to BCD, B \to ACD\}$
>
> $G_2 = \{A \to BE, B \to AE\}$
>
> $G_3 = \{B \to HC\}$

The interrelational fd's are as follows:

> $G(i,i) = G_i, \ 1 \le i \le 3$
>
> $G(1,2) = G(2,1) = \{A \to B, B \to A\}$
>
> $G(1,3) = G(3,1) = \{B \to C\}$
>
> $G(2,3) = G(3,2) = \emptyset$

For every fd $X \to Y \in G_i, 1 \le i \le 3$, $X^*_{w,i} = X$, so $Y - X^*_{w,i} = Y$. It can be verified that for each i, $weak - unique(i) = \mathbf{true}$. So $\mathbf{R}$ is weakly independent wrt G. □

5 Polynomial Time Conditions

In this section, we give two polynomial time conditions, one is sufficient and the other is necessary, for weak independence. Both conditions are general enough to test all independent schemes as special case.

Theorem 3 (Sufficient Condition) *The following statements are equivalent and sufficient for R_i to satisfy the weak uniqueness condition wrt G*

1. *For every fd $X \to Y \in G_j$ with $R_j \neq R_i$ and every attribute $A \in Y - X_{w,j}^*$, if a sequence $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ is a minimal derivation of XA from R_i wrt $G(X, R_j)$, then either $XA \subseteq R_i$ or $XA \subseteq X_k Y_k$ for some $1 \leq k \leq l$.*

2. *(Polynomial time test) For every fd $X \to Y \in G_j$ with $R_j \neq R_i$ and every attribute $A \in Y - X_{w,j}^*$, either $XA \subseteq R_i$ or $(R_i)_\alpha^+ \not\supseteq XA$, where α is the set of fd's in $G(X, R_j)$ in which $X \to A$ is not embedded.*

Proof: First, note that statement 2 is essentially a restatement of statement 1, so they are equivalent. For the sufficiency, it can be seen that any of $XA \subseteq R_i$ and $XA \subseteq X_k Y_k$ implies $G(\sigma) \models X \to A$. $\square$

It is easy to see that the condition in Theorem 3 is not necessary. Consider the case in Example 6 where $i = 2$, $j = 1$ and $X \to Y$ is $A \to BCD \in G_j$. $B \to HC$ is a minimal derivation of AC from R_i wrt $G(A, R_j)$, but $AC \not\subseteq R_i$ and $AC \not\subseteq BHC$. However, there is an fd $Q \to W$ equivalent to $A \to BCD$ and some attribute $I \in W - X_{w,j}^*$ such that either $QI \subseteq R_i$ or $QI \subseteq BHC$. In particular, $B \to ACD$ is such an fd with $I = C$. This situation is summarized in the following theorem which gives a polynomial time necessary condition for the weak uniqueness. Recall that, for any lhs X of G_j, $EQ(X, R_j)$ denotes the set of fd's in G_j having lhs's equivalent to X.

Theorem 4 (Necessary Condition) *The following two statements are equivalent and necessary for R_i to satisfy the weak uniqueness condition wrt G.*

1. *For every $X \to Y \in G_j$ with $R_i \neq R_j$ and every attribute $A \in Y - X_{w,j}^*$, if a sequence $X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ is a minimal derivation of XA from R_i wrt $G(X, R_j)$, then there exists an fd $Q \to W \in EQ(X, R_j)$ and an attribute $B \in W - X_{w,j}^*$ such that $Q \subseteq X_{w,j}^*$ and either $QB \subseteq R_i$ or $QB \subseteq X_k Y_k$ for some $1 \leq k \leq l$.*

2. *(Polynomial time test) For every fd $X \to Y \in G_j$ with $R_j \neq R_i$ and every attribute $A \in Y - X_{w,j}^*$, if $(R_i)_{G(X,R_j)}^+ \supseteq XA$, then there exists an fd $Q \to W \in EQ(X, R_j)$ and an*

attribute $B \in W - X^*_{w,j}$ such that $Q \subseteq X^*_{w,j}$ and either $QB \subseteq R_i$ or $(R_i)^+_\alpha \not\supseteq XA$, where α is the set of fd's in $G(X, R_j)$ in which $Q \to B$ is not embedded.

Proof: As in Theorem 3, statement 2 is just a restatement of statement 1, so they are equivalent.

Assume that R_i satisfies the weak uniqueness condition wrt G. We show statement 1 holds. Let $X \to Y \in G_j$ with $R_j \neq R_i$ and $A \in Y - X^*_{w,j}$ such that a sequence $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ is a minimal derivation of XA from R_i wrt $G(X, R_j)$. We may assume $XA \not\subseteq R_i$ and therefore $l \geq 1$; otherwise, by letting $Q \to W$ be $X \to Y$ and B be A, statement 1 holds trivially. We construct a state $\rho = \{v_i, u_1, \ldots, u_l, v_j\}$ based on derivation σ exactly as in the proof of Lemma 5, except that v_j has constant 0's exactly in columns $X^*_{w,j}$ rather than in $X^+_{G(\sigma)}$. $v_j[A]$ is a distinct constant because $A \notin X^*_{w,j}$. By $G(\sigma) \models X \to A$ and Lemma 2, ρ is not embedded-satisfying G. Then from Lemma 4, there are the following cases of violation of embedded-satisfaction:

Case 1. v_i and v_j violate embedded-satisfaction. That is, $\{v_i[R_i \cap R_j], v_j[R_i \cap R_j]\} \subseteq \rho(i,j)$ and there exists an fd $Q \to Z \in G(i,j)$ such that $v_i[Q] = v_j[Q]$ and $v_i[B] \neq v_j[B]$ for some $B \in Z$. Then $B \notin X^*_{w,j}$ and $Q \subseteq X^*_{w,j}$ from the construction of ρ. Let $Q \to W$ be the fd in G_j such that $B \in W$. Since $Q \subseteq XY$ and XY is locally closed wrt G, $QW \subseteq XY$ and so $Q \to W$ is weaker than $X \to Y$. But since $W \not\subseteq X^*_{w,j}$ (because $B \in W - X^*_{w,j}$), $Q \to W$ is not strictly weaker than $X \to Y$, therefore, $Q \to W \in EQ(X, R_j)$. We have shown that $QB \subseteq R_i$ for some fd $Q \to W \in EQ(X, R_j)$ such that $Q \subseteq X^*_{w,j}$ and $B \in W - X^*_{w,j}$. Thus the statement 1 holds.

Case 2. u_{q_k} and v_j violate embedded-satisfaction for some $1 \leq q_k \leq k$. There exists an fd $Q \to Z \in G(q_k, j)$ such that $u_{q_k}[Q] = v_j[Q]$ and $u_{q_k}[B] \neq v_j[B]$ for some $B \in Z$. Note that $QB \subseteq X_k Y_k$ and $Q \subseteq X^*_{w,j}$ and $B \notin X^*_{w,j}$ from the construction of ρ and local closure of $X_k Y_k$. Let $Q \to W \in G_j$ such that $B \in W$. Exactly as in Case 1, we can show $Q \to W$ and $X \to Y$ are equivalent. So we have shown that $QB \subseteq X_k Y_k$ for an fd $Q \to W \in EQ(X, R_j)$ such that $Q \subseteq X^*_{w,j}$ and $B \in W - X^*_{w,j}$. Thus statement 1 holds.□

It is easy to see that when $G(i,j) = \emptyset$ for every $i \neq j$, conditions in both Theorems 3 and 4 are reduced to the uniqueness condition for independent scheme [23, 24]: for every $X \to Y \in G_j$ with $R_i \neq R_j$ and every attribute $A \in Y - X^*_{w,j}$, $(R_i)^+_{G(X,R_j)} \not\supseteq XA$. The next theorem gives a more general condition on which this reduction happens.

Theorem 5 *Assume that, for every $X \to Y \in G_j$ and every $Q \to W \in EQ(X, R_j)$ such that $Q \subseteq X^*_{w,j}$, the following conditions hold:*

1. $QB \not\subseteq R_i$ *for every $R_i \neq R_j$ and every $B \in W - X^*_{w,j}$; and*

2. $QB \not\subseteq UV$ *for every $U \to V \in G(X, R_j)$ and every $B \in W - X^*_{w,j}$.*

(However, it is possible that $QB \subseteq R_i$ or $QB \subseteq UV$ for attributes B in $W \cap X_{w,j}^*$) Then (a) the sufficient condition in Theorem 3 and the necessary condition in Theorem 4 are equivalent to the uniqueness condition of [23, 24]; (b) **R** *is independent wrt G if and only if* **R** *is weakly independent wrt G.*

Note that when $G(i,j) = \emptyset$ for every $i \neq j$, conditions 1 and 2 in Theorem 5 automatically hold. In fact, for any $U \to V \in G_j$, $QB \subseteq UV$ implies that $Q \to W$ is weaker than $U \to V$, therefore $U \to V \notin G(X, R_j)$.

Proof of Theorem 5: Part (a) follows from the assumption of the theorem and the first statements in both (polynomial time) conditions. To see Part (a), clearly, every independent scheme is weakly independent . Assume **R** is weakly independent wrt G. Then the necessary condition in Theorem 4 holds for every R_i. Then from Part (a), the uniqueness condition holds for every R_i, so **R** is independent wrt G. $\square$

6   Weak Independence and Acyclicity

In this section, we show that a certain acyclicity of database schemes implies weak independence. When no fd is embedded in two relation schemes, this acyclicity implies independence. Therefore, acyclicity of database schemes is also desirable in constraint enforcement.

A *hypergraph* $\mathcal{H}$ is a pair $(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is a set *nodes* and $\mathcal{E}$ is a set of *edges* which are arbitrary non-empty subsets of $\mathcal{N}$. A *path* from node s to node t is a sequence of $k \geq 1$ of edges $(e_1, \ldots, e_k)$ such that

1. s is in e_1,

2. t is in e_k, and

3. $e_i \cap e_{i+1}$ is non-empty if $1 \leq i < k$.

We also say that the above sequence of edges is a path from e_1 to e_k.

A *weak β-cycle* [11] in a hypergraph $\mathcal{H}$ is a sequence $(S_1, x_1, S_2, x_2, \ldots, S_m, x_m, S_{m+1})$ such that

1. $x_1, \ldots, x_m$ are distinct nodes of $\mathcal{H}$;

2. $S_1, \ldots, S_m$ are distinct edges of $\mathcal{H}$, and $S_{m+1} = S_1$;

3. $m \geq 3$, that is, there are at least 3 edges involved; and

4. x_i is in S_i and S_{i+1} $(1 \leq i \leq m)$ and in no other S_j.

It is sometimes convenient to refer to the sequence $(S_1, \ldots, S_m, S_{m+1})$ of edges alone of a weak β-cycle as a weak β-cycle. A hypergraph is said to be β-*acyclic* if it has no weak β-cycle [11].

A γ-*cycle* in a hypergraph $\mathcal{H}$ is a sequence $(S_1, x_1, S_2, x_2, \ldots, S_m, x_m, S_{m+1})$ such that

1. $x_1, \ldots, x_m$ are distinct nodes of $\mathcal{H}$;

2. $S_1, \ldots, S_m$ are distinct edges of $\mathcal{H}$, and $S_{m+1} = S_1$;

3. $m \geq 3$, that is, there are at least 3 edges involved;

4. x_i is in S_i and S_{i+1} $(1 \leq i \leq m)$; and

5. if $1 \leq i < m$, then x_i is in no S_j except S_i and S_{i+1}.

Note that the only difference between a γ-cycle and a weak β-cycle is that "$1 \leq i < m$" in (5) is replaced by "$1 \leq i \leq m$" to define a weak β-cycle. Thus every weak β-cycle is a γ-cycle. A hypergraph is said to be γ-*acyclic* if it has no γ-cycle. If a hypergraph is γ-acyclic then it is also β-acyclic [11].

The hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ defined by a database scheme $\mathbf{R}$ is such that $\mathcal{N} = \cup \mathbf{R}$ and $\mathcal{E} = \mathbf{R}$. We say a database scheme $\mathbf{R}$ is θ-acyclic, for $\theta = \beta, \gamma$, if its hypergraph is θ-acyclic. Let $\mathbf{R} = \{R_1, \ldots, R_m\}$ and $G = G_1 \cup \ldots \cup G_m$ be a set of fd's in the form assumed in Section 3. A *derivation hypergraph* wrt R_i and G, denoted $\mathcal{H}_i^G = (\mathcal{N}, \mathcal{E})$, is a hypergraph such that $\mathcal{N} = \cup \mathbf{R}$ and $\mathcal{E}$ is given by

$$\mathcal{E} = \{e \mid e = XY, X \to Y \in G\} \cup \{R_i\}.$$

Theorem 6 *If $\mathcal{H}_i^G$ is β-acyclic and contains no γ-cycle of length 3, then R_i satisfies weak uniqueness condition wrt G.*

Proof: Assume that R_i does not satisfy the weak uniqueness condition wrt G. Then for some fd $X \to Y \in G_j$ with $R_j \neq R_i$ and some attribute $A \in Y - X_{w,j}^*$, there is a derivation $\sigma : X_1 \to Y_1 \in G_{q_1}, \ldots, X_l \to Y_l \in G_{q_l}$ of XA from R_i wrt $G(X, R_j)$. Since $G(\sigma) \not\models X \to A$, we have $XA \not\subseteq R_i$ and $XA \not\subseteq X_k Y_k$, for $1 \leq k \leq l$. There must exist two attributes B and C in XY and two edges e_B and e_C in $\{R_i, X_1 Y_1, \ldots, X_l Y_l\} \subseteq \mathcal{E}$, where $\mathcal{H}_i^G = (\mathcal{N}, \mathcal{E})$, such that $B \in e_B - e_C$ and $C \in e_C - e_B$. This is guaranteed because XA is not contained in any $X_k Y_k$ or in R_i. Clearly, there is a path from B to C through edges in $\{R_i, X_1 Y_1, \ldots, X_l Y_l\}$. We assume that the above choices of B, C, e_B, and e_C are made so that this path is as short as possible. (It should be emphasized that the minimality of the path length has also considered the choice of attributes B and C.) Let this path be

$$(e_1, \ldots, e_n), \quad n \geq 2,$$

where $e_1 = e_B$ and $e_n = e_C$. Then by the minimality of the length of the path $(e_1, \ldots, e_n)$, we can show that the path

$$(e_n, C, XY, B, e_1, x_1, e_2, x_2, \ldots, e_{n-1}, x_{n-1}, e_n)$$

is either a γ-cycle of length 3 when $n = 2$, or a weak β-cycle when $n > 2$, where x_i is any attribute in $e_i \cap e_{i+1}$; that is, except for the case of $n = 2$ and $i = n - 1$, any of C, B and x_i appear only in its two adjacent edges, no where else. Consider the following cases.

Case 1. $n = 2$. Then the above path becomes

$$(e_2, C, XY, B, e_1, x_1, e_2).$$

Clearly, $B \notin e_2$ since $e_2 = e_C$. Also, $C \notin e_1$ since $e_1 = e_B$. This means that the path is a γ-cycle in $\mathcal{H}_i^G$.

Case 2. $n > 2$. Clearly, $B \notin e_i$, $2 \leq i \leq n$; otherwise, $(e_1, \ldots, e_n)$ is not the shortest path from B to C. Similarly, $C \notin e_i$, $1 \leq i \leq n - 1$. Now we show that x_i is in no e_j except e_i and e_{i+1}. Assume $x_i \in e_j$ for some $j > i + 1$. Then $x_i \in e_i \cap e_j$, and we can replace path $(e_1, \ldots, e_n)$ by $(e_1, \ldots, e_i, e_j, \ldots, e_n)$, which is strictly shorter than $(e_1, \ldots, e_n)$, a contradiction. Similarly, if $x_i \in e_j$ for some $j < i$, then $x_i \in e_j \cap e_{i+1}$ and we can replace $(e_1, \ldots, e_n)$ by a shorter path $(e_1, \ldots, e_j, e_{i+1}, \ldots, e_n)$, a contradiction again. To complete the proof of γ-acyclicity, we must show x_i is not in XY either. Suppose $x_i \in XY$. Clearly, $x_i \neq B$ and $x_i \neq C$; otherwise, $(e_1, \ldots, e_n)$ is not the shortest path from B to C. Consider the following subcases:

Case 2.1. $i < n - 1$. From the above results, $x_i \in e_i - e_n$ and $C \in e_n - e_i$. Thus we can choose x_i as B in the first place and have a shorter path $(e_i, \ldots, e_n)$, a contradiction.

Case 2.2. $i = n - 1$. Similarly, from the above results, we have $x_i \in e_i - e_1$ and $B \in e_1 - e_i$. Therefore, we can choose x_i as C and have a shorter path $(e_1, \ldots, e_i)$, a contradiction.

In all cases, we have deduced a contradiction, so R_i satisfies the weak uniqueness condition wrt G. $\square$

However, the condition that $\mathcal{H}_i^G$ is β-acyclic and contains no γ-cycle of length 3 is not necessary for R_i to satisfy the weak uniqueness condition. Consider the weakly independent scheme in Example 6, we have $\mathcal{H}_1^G = (\{ABCDEH\}, \{ABCD, ABE, BHC\})$. Clearly, $\mathcal{H}_1^G$ is β-acyclic but contains a γ-cycle of length 3: $(ABCD, C, BHC, B, ABE)$.

Corollary 1 *If $\mathcal{H}_i^G$ is γ-acyclic for every $R_i \in \mathbf{R}$, then $\mathbf{R}$ is weakly independent wrt G.*

Proof: This follows immediately from Theorems 1 and 6 and γ-acyclicity $\Rightarrow$ β-acyclicity.$\square$

Corollary 2 *If $G(i, j) \neq \emptyset$ for some $i \neq j$, then $\mathbf{R}$ is not independent wrt G; else, if $\mathcal{H}_i^G$ is γ-acyclic for every $R_i \in \mathbf{R}$, then $\mathbf{R}$ is independent wrt G.*

Proof: The first part follows from a result in [13]; the second part follows from Proposition 1 and

Corollary 1.□

Corollary 3 *Assume* $G = G_1 \cup \ldots \cup G_m$ *is a set of key dependencies; that is, for every fd* $X \to Y \in G_i$, $R_i = XY$. *If* **R** *is* β-*acyclic and contains no* γ-*cycle of length 3, then* **R** *is weakly independent wrt* G.

Proof: First note that the derivation hypergraph $\mathcal{H}_i^G$ becomes a set of relation schemes of **R** because XY is a relation scheme for every fd $X \to Y \in G$. Therefore, $\mathcal{H}_i^G$ is a subhypergraph of the hypergraph of **R**. Since subhypergraph preserves β- and γ-acyclicity [11], each $\mathcal{H}_i^G$ must be β-acyclic and contains no γ-cycle of length 3. Then by Theorems 1 and 6, the corollary follows.□

It was shown in [9] that BCNF and γ-acyclicity together imply a property similar to weak independence (but the notion of weak independence was not defined there). Corollary 3, which assumes no BCNF or γ-acyclicity, is a generalization of that result.

7 Extension to Fd's and Jd $\bowtie$ **R**

The *join dependency (jd)* $\bowtie$ **R** is satisfied by a relation I over $\cup$**R** if $\prod_{R_1}(I) \bowtie \ldots \bowtie \prod_{R_m}(I) = I$, where **R** $= \{R_1, \ldots, R_m\}$. The database scheme is called *lossless* [2] when the jd is implied by its dependencies. Let $(\mathbf{R}, G \cup \{\bowtie \mathbf{R}\})$ be a database scheme, where G is a set of embedded fd's as assumed in Section 3. From a result in [13], G is a cover of all fd's implied by $G \cup \{\bowtie \mathbf{R}\}$. So we have the following definition.

Definition 2 **R** is said to be *weakly independent* wrt $G \cup \{\bowtie \mathbf{R}\}$ if every embedded-satisfying state wrt G is consistent wrt $G \cup \{\bowtie \mathbf{R}\}$.□

The following theorem reduces the test of weak independence wrt $G \cup \{\bowtie \mathbf{R}\}$ to the case wrt G alone.

Theorem 7 **R** *is weakly independent wrt* $G \cup \{\bowtie \mathbf{R}\}$ *if and only if* **R** *is weakly independent wrt* G.

Proof: This follows from definitions and that I is a weak instance of a state ρ wrt G implies that $\prod_{R_1}(I) \bowtie \ldots \bowtie \prod_{R_m}(I)$ is a weak instance of ρ wrt $G \cup \{\bowtie \mathbf{R}\}$.□

8 Open Problems

Three questions remain to be answered: (a) Is there a polynomial time test for weakly independent schemes? if so, what is it? (b) Can a similar result regarding polynomial time construction of relational expressions for computing *total projections* as for independent schemes [22, 16, 3] be found for weakly independent schemes? Such a result is relevant in efficient query answering

under the *representative instance approach* [17, 22]. From a result in [15], such expressions can be constructed in exponential time for weakly independent schemes. (c) When and how can a weakly independent scheme be designed to embed a given set of functional dependencies?

References

[1] W.W. Armstrong: "Dependency structures of data base relationships", *IFIP Congress*, 1974, pp. 580-583

[2] A.V. Aho, C. Beeri, and J.D. Ullman: The theory of joins in relational databases, *ACM TODS*, Vol. 4, No. 3, 1979, pp. 297-314

[3] P. Atzeni and E.P.F. Chan: "Efficient query answering in the representative instance approach", in proceedings of *ACM Symposium on Principles of Database Systems*, 1985, pp. 181-188

[4] C. Beeri and P.A. Bernstein: "Computational problems related to the design of normal form relational schemas", *ACM TODS*, Vol. 4, No. 1, 1979, pp. 30-59

[5] C. Beeri, P.A. Bernstein, and N. Goodman: "A sophisticate's introduction to database normalization theory", in proceedings of *International Conference on Very Large Data Bases*, 1978, pp. 113-124

[6] J. Biskup, U. Dayal, and P.A. Bernstein: "Synthesizing independent database schemes", in proceedings of *International Conference on Management of Data*, 1979, pp. 143-151

[7] C. Beeri, R. Fagin, D. Maier, A.O. Mendelson, J.D. Ullman, and M. Yannakakis: "Properties of acyclic database schemes", in proceedings of *ACM Symposium on Theory of Computing*, 1981, pp. 355-362

[8] C. Beeri and J. Rissanen: "Faithful representation of relational database schemes", IBM, RJ2722, 1980

[9] E.P.F. Chan and H.J. Hernandez: "On the desirability of γ-acyclic BCNF database schemes", *Theoretical Computer Science*, No. 62, 1988, pp. 67-104

[10]   E.F. Codd: "A relational model for large shared data banks", *CACM*, Vol. 13, 1970, pp. 377-387

[11]   R. Fagin: "Degrees of Acyclicity for hypergraphs and relational database schemes", *JACM*, Vol. 30, No., 3, July 1983, pp. 514-550

[12]   M.H. Graham and K. Wang: "Constant time maintenance or the triumph of fd's", in proceedings of *ACM Symposium on Principles of Database Systems*, 1986, pp. 202-216

[13]   M.H. Graham and M.Yannakakis: "Independent database schemes", *JCSS*, Vol. 28, No. 1, 1984, pp. 121-141

[14]   P. Honeyman: "Test satisfaction of functional dependencies", *JACM*, Vol. 29, No. 3, (1982), pp. 668-677

[15]   H.J. Hernandez and K. Wang: "On the boundedness of constant-time-maintainable database schemes", *SIAM Journal on Computing*, Vol. 22, No. 1, Feb. 1993, pp. 29-45

[16]   M. Ito, M. Iwasaki and T. Kasam: "Some results on the representative instances in relational databases", *SIAM Journal on Computing*, Vol. 14, No. 2, May, 1985, pp. 334-353

[17]   A.O. Mendelzon: "Database states and their tableaux", *ACM TODS*, Vol. 9, No. 2, 1984, pp. 264-282

[18]   D. Maier: The theory of relational databases, CSP Inc., 1983

[19]   D. Maier, J.D. Ullman, and M.Y. Vardi: "On the foundations of the universal relational model", *ACM TODS*, Vol. 9, No. 2, 1984, pp. 283-308

[20]   D. Maier, A.O. Mendelzon, Y. Sagiv: "Testing implication of data dependencies", *ACM TODS*, Vol. 2, No. 3, 1977, pp. 455-469

[21]   Y. Sagiv: "Can we use the universal instance assumption without using nulls?" in proceedings of *International Conference on Management of Data*, 1981, pp. 108-120

[22]   Y. Sagiv: "A characterization of globally consistent databases and their correct access paths", *ACM TODS*, Vol. 8, No. 2, (1983), pp. 266-286

[23]   Y. Sagiv: "On computing restricted projections of representative instances", in proceedings of *ACM symposium on Principles of Database Systems*, 1985, pp. 171-180

[24] Y. Sagiv: "Evaluation of queries in independent database schemes", *JACM*, Vol. 38, No. 1, 1991, pp. 120-161

[25] J.D. Ullman: Principles of database systems, CSP, Inc., 1982

[26] Y. Vassiliou: "A formal treatment of imperfect information in database management", Ph.D thesis, University of Toronto, 1980

[27] K. Wang: "Can constant-time-maintainability be more practical?" in proceedings of *ACM Symposium on Principles of Database Systems*, 1989, pp. 120-127

[28] M.Yannakakis: "Algorithms for acyclic database schemes", in proceedings of *International Conference on Very Large Data Bases*, 1981, pp. 82-94