WEAK MULTIVALUED DEPENDENCIES+

Patrick C. Fischer
Dirk Van Gucht

Vanderbilt University

1. Introduction. Weak multivalued dependencies (WMVDs) were introduced by Jaeschke and Scheck in order to characterize when two NEST operations on different single attributes would commute [JS]. Thomas extended this result to nesting on arbitrary structures and also gave a graph-theoretic characterization [T].

In this paper we give additional characterizations of WMVDs. We also give an axiom system for WMVDs and prove its completeness. We then extend results of [JS] and [T] to more than two NEST operations and show that the satisfaction of a certain set of WMVDs yields a reasonably natural horizontal and vertical decomposition of a relation even when the corresponding regular MVD is not satisfied. Other applications of WMVDs are discussed.

Finally, we contrast WMVDs with the degenerate MVDs of Armstrong and Delobel [AD] and the different version of degenerate MVDs given by Kambayashi and Yoshikawa [KY]. We also propose an axiom system to handle the interactions between WMVDs and ordinary MVDs, but its completeness has not been established.

2. Weak Multivalued Dependencies. We define WMVDs as a particular subclass of the template dependencies (TDs) [BV, SU1, FMUY] and will show this definition equivalent to the one given in [JS].

Definition 1. Let U be a set of attributes and let X, Y, Z be subsets of U such that Z = U-XY. A weak multivalued dependency, written X -w-> Y, is a template dependency consisting of a hypothesis of three rows t_1, t_2, t_3 and a conclusion row t_4 such that

1. $t_1[X] = t_2[X] = t_3[X] = t_4[X]$
2. $t_1[Y] = t_2[Y]$
3. $t_1[Z] = t_3[Z]$
4. $t_4[Y] = t_3[Y]$
5. $t_4[Z] = t_2[Z]$

In tableau form, X -w-> Y is the WMVD

	X	Y-X	Z
t_1:	x	y	z
t_2:	x	y	z'
t_3:	x	y'	z
t_4:	x	y'	z'

A relation (instance) r over U satisfies X -w-> Y if r satisfies the TD $(t_1, t_2, t_3)/t_4$ given above. (In contrast, the ordinary MVD X ->-> Y would correspond to the TD $(t_2, t_3)/t_4$).

Remark 1. The roles of t_1, t_2, t_3 are not symmetric. The tuple t_1 must satisfy

$$t_1[A]=t_2[A] \vee t_1[A]=t_3[A]$$

for each attribute $A \in U$. Then, t_2 is the tuple agreeing with t_1 on Y and t_3 is the tuple agreeing with t_1 on Z. Clearly if a relation satisfies X ->-> Y, then it satisfies X -w-> Y. The following example shows the converse is not true.

+ This research was partially supported by NSF Grant MCS-8115106.

	A	B	C
s_1:	a	b_1	c_1
s_2:	a	b_1	c_2
s_3:	a	b_2	c_1
s_4:	a	b_2	c_2
s_5:	a	b_3	c_3

This relation r_{ex} satisfies A -w-> B, but does not satisfy A ->-> B. The same is true of the relation consisting only of s_1 and s_4. On the other hand, the relation obtained by deleting s_4 from r_{ex} does not satisfy A -w-> B since the first three tuples fit the hypothesis of the TD and the deleted tuple fits the conclusion.

We first give several alternate characteristics of the notion of a WMVD.

Theorem 1. Let U be a relation scheme, X, Y subsets of U, Z = U-XY, and let r be a relation over U. Then the following are equivalent:

1. X -w-> Y holds in r (i.e., is satisfied by r).
2. Let $Y(x,z)=\{t[Y] \mid t \in r \ \& \ t[X]=x \ \& \ t[Z]=z\}$.

   Then for any tuples $t_1, t_2 \in r$, if $t_1[XY] = t_2[XY]$ then $Y(x, t_1[Z]) = Y(x, t_2[Z])$. (This is essentially the definition given in [JS] for a WMVD.)
3. Let G(x) be a bipartite graph (bigraph) constructed as follows:

   a. Let $r(x) = \{t \in r \mid t[X] = x\} = \sigma_{X=x}(r)$.

   b. The first set of nodes of G(x) is $\Pi_Y(r(x))$ and the second set of nodes of G(x) is $\Pi_Z(r(x))$.

   c. The edges of G(x) connect t[Y] to t[Z] for each $t \in r(x)$.

   Then for each $x \in \Pi_X(r)$, each connected component of G(x) is a complete bigraph (cf. [T].)
4. The relation r can be partitioned horizontally into subrelations $r_1, r_2, \ldots, r_n$ such that

   a. r_i satisfies the MVD X ->-> Y for $1 \leq i \leq n$;

   b. $\Pi_{XY}(r_i) \cap \Pi_{XY}(r_j) = \emptyset$ for $1 \leq i < j \leq n$;

   c. $\Pi_{XZ}(r_i) \cap \Pi_{XZ}(r_j) = \emptyset$ for $1 \leq i < j \leq n$.

Proof.

(1 => 2) We have to prove that each element y of $Y(x,t_1[Z])$ belongs to $Y(x,t_2[Z])$.

Since $y \in Y(x,t_1[Z])$ we know that r contains a tuple such that $t_3[X]=x$, $t_3[Y]=y$, and $t_3[Z]=t_1[Z]$.

Since X -w-> Y holds r contains a tuple t_4 with $t_4[X]=x$, $t_4[Y]=y$, and $t_4[Z]=t_2[Z]$.

Hence $y \in Y(x,t_2[Z])$. The reverse inclusion follows by symmetry.

(2 => 3) This was shown in [T].

(3 => 4) For each $x \in \Pi_X(r)$ consider the bipartite graph G(x). G(x) has k_x connected components, each being a complete bigraph.

Let $G_i(x)$ be the i-th connected component of G(x). We associate with $G_i(x)$ the subrelation $r_i(x)$ consisting of all tuples of the form (x,y,z) where y and z are nodes of $G_i(x)$. By the definition of G(x) and the fact that $G_i(x)$ is a complete bigraph we know that $r_i(x)$ forms a subrelation of r satisfying the MVD X ->-> Y.

Consider the family of relations
$R=\{r_i(x) \mid x \in \Pi_X(r) \text{ and } 1 \leq i \leq k_x\}$.

Clearly R forms a (horizontal) partition of r. Now consider two different elements $r_i(x)$ and $r_j(x')$ of R. If $x \neq x'$ then conditions 4(b) and 4(c) are trivially satisfied. If x=x' then conditions 4(b) and 4(c) follow from the fact that $r_i(x)$ and $r_j(x')$ correspond to two different connected components of G(x).

(4 => 1) Suppose the WMVD X -w-> Y is not satisfied. Then r contains tuples of the form $t_1=\langle x,y,z\rangle$, $t_2=\langle x,y,z'\rangle$ and $t_3=\langle x,y',z\rangle$ but not the tuple $\langle x,y',z'\rangle$. Let r_i be the subrelation containing t_1. Since t_2 and t_3 have the same XY projection and t_1 and t_3 the same XZ projection it follows from conditions 4(b) and 4(c) that t_2 and t_3 also belong to r_i. From 4(a) it follows that $\langle x, y',z'\rangle$ is an element of r_i, hence an element of r, a contradiction.

3. A Complete Axiom System for Weak MVDs.

We give the following axiom system (really a set of inference rules) for WMVDs. Let U be a set of attributes and let W,X,Y be subsets of U.

W1. Reflexivity. X -w-> X.

W2. Transport Rule. If X -w-> Y and $V \subseteq X$, then X -w-> Y-V and X -w-> YV.

W3. Augmentation. If X -w-> Y, then XW -w-> Y.

W4. Complementation. If X -w-> Y, then X -w-> U-XY.

The transport rule must be explicitly stated here since there is no transitivity axiom for WMVDs. Clearly, the addition of the appropriate transitivity axiom would yield an axiom system for MVDs equivalent to that in [B].

If one compares our axioms with those in [A] one can see that Armstrong's reflexivity axiom is really a blending of the forms of W1 and W2 and his augmentation axiom is really a blending of the forms of W3 and W2. While we have been forced to state the transport rule explicitly by the special characteristics of WMVDs we feel that the understanding of the axiom systems for functional dependencies (FDs) as given in [A] and of MVDs as given in [B] could be sharpened by explicit use of the transport rule in these systems.

In the proof of the completeness of W1 through W4 we use techniques similar to those in [SU1].

Lemma 1. Let D be a set of full template dependencies and let d be a full TD. D logically implies d if and only if chasing the tableau consisting of the hypothesis of d causes us to generate the conclusion t of d.

Proof. This is a special case of a lemma in [SU1], also in [BV].

Lemma 2. Let H be a tableau consisting of the strings t_1,t_2,t_3. Let X -w-> Y be a WMVD and assume t_4 is generated by applying the TD d corresponding to X -w-> Y once to H. Then the TD $(t_1,t_2,t_3)/t_4$ is equivalent to a family of WMVDs, each of which can be inferred from X -w-> Y using rules W2, W3, and W4.

Proof.

From axiom W2, we may assume $X \cap Y = \emptyset$ without loss of generality, hence X, Y and Z=U-XY are pairwise disjoint.

The template dependency $d=(s_1,s_2,s_3)/s_4$ corresponding to X -w-> Y has the following format:

	X	Y	Z
s_1:	x	y	z
s_2:	x	y	z'
s_3:	x	y'	z
s_4:	x	y'	z'

First, we derive the following subsets of U from the tableau $H=\{t_1,t_2,t_3\}$:

$$X = \{A \in U \mid t_1[A]=t_2[A]=t_3[A]\}$$
$$V = X'-X$$
$$Y_1 = Y \cap V$$
$$Z_1 = Z \cap V$$
$$Y_2 = Y-Y_1$$
$$Z_2 = Z-Z_1$$

Hence, $V=Y_1Y_2$, $Y=Y_1Y_2$, $Z=Z_1Z_2$. Also, Y_1,Y_2,Z_1,Z_2 and, X are pairwise disjoint (some of them possibly empty).

Since d applied to H yields a tuple t_4, we know from the definition of the chase that there must exist an ordering of the tuples t_1,t_2,t_3, say t_{i_1},t_{i_2},t_{i_3} (where (i_1,i_2,i_3) is a permutation of the integers 1,2,3) such that

a. if s_j and s_k agree on an attribute A then so do t_{i_j} and t_{i_k} for $1 \le j<k \le 3$.

b. if s_j and s_4 agree on an attribute A then so do t_{i_j} and t_4 for $1 \le j \le 3$.

(The ordering for i_1,i_2,i_3 could be determined explicitly using the principles of Remark 1 above.) Hence $X \subseteq X'$ and $X'=XV$.

For simplicity we assume that t_1,t_2,t_3 are in the desired order. Then the TD $(t_1,t_2,t_3)/t_4$ has the following format:

	X	Y_1	Z_1	Y_2	Z_2
t_1:	x	y_1	z_1	y_2	z_2
t_2:	x	y_1	z_1	y_2	z'_2
t_3:	x	y_1	z_1	y'_2	z_2
t_4:	x	y_1	z_1	y'_2	z'_2

It follows from the definition of WMVDs that $(t_1,t_2,t_3)/t_4$ is equivalent to the family of WMVDs X' -w-> Y' where

X' is defined as above (i.e., $X'=XY_1Z_1=XV$)

Y' satisfies $Y_2 \subseteq Y' \subseteq U-Z_2$ or $Z_2 \subseteq Y' \subseteq U-Y_2$

So what is left to prove is that each member X' -w-> Y' of the family of WMVDs can be inferred from X -w-> Y using rules W_2,W_3,W_4. By augmentation, rule W3, we know that X -w-> Y implies XV = X' -w-> Y. By the transport rule W2,

$X' \text{-w->} Y_2$ follows since $Y_1=(Y-Y_2) \subseteq X'$.
We have two cases:

<u>Case 1</u> $Y_2 \subseteq Y' \subseteq U-Z_2$

Clearly $Y'-Y_2 \subseteq X'$. Thus $X' \text{-w->} Y'$ follows from $X' \text{-w->} Y_2$ by rule W_2.

<u>Case 2</u> $Z_2 \subseteq Y' \subseteq U-Y_1$

We apply the complementation rule W_4 to $X' \text{-w->} Y_2$ and conclude that $X' \text{-w->} Z_2$ must hold. Since $Y'-Z_2 \subseteq X'$, $X' \text{-w->} Y'$ follows using rule W_2.

The next lemma shows that different WMVDs interact only in a fairly simple way.

<u>Lemma 3</u>. Let H be a tableau consisting of distinct strings t_1,t_2,t_3.

Let D be a set of TDs corresponding to a set of WMVDs. Let t_4 be a string generated in one step by chasing H under D. Let $H' = \{t_1,t_2,t_3,t_4\}$. Then chasing H' under D yields H', i.e., no new strings can be obtained from H' by using any of the TDs in D. Hence H' is the chase of H under D.

<u>Proof</u>. Let $d \in D$ be a WMVD generating t_4. From Remark 1, we know that there exists an ordering t_{i_1},t_{i_2},t_{i_3} of the tuples t_1,t_2,t_3 such that H has the following format:

	X	Y	Z
t_{i_1}:	x	y	z
t_{i_2}:	x	y	z'
t_{i_3}:	x	y'	z

Where $X=\{A \mid t_1[A]=t_2[A]=t_3[A]\}$

$$Y=\{A \mid t_{i_1}[A]=t_{i_2}[A]\}-X$$

$$Z = \{A \mid t_{i_1}[A]=t_{i_3}[A]\} - X$$

From the definition of X,Y,Z it follows that y and y' disagree on each attribute in Y. The same holds for z and z' over Z. Also X,Y,Z are clearly pairwise disjoint and nonempty (t_1,t_2,t_3 are distinct strings). Furthermore, XYZ = U, else no WMVD could be applied to H.

Since d is applicable to H and d is a WMVD, d has to generate the tuple $\langle x,y',z'\rangle$. Note also that if $d' \in D$ is applicable to H it has to generate the same tuple. So $t_4=\langle x,y',z'\rangle$ is the only tuple that can be generated from H in one step using the dependencies in D.

The reader can easily verify that if we take an arbitrary 3-element subset S of H' and apply D to this subset that the only tuple we can generate is the element of H' not belonging to S. Hence H' is the chase of H under D.

<u>Theorem 2</u>. Rules W1, W2, W3, W4 form a complete axiom system for WMVDs.

<u>Proof</u>. Assume $X \text{-w->} Y$ is a WMVD d logically implied by a set D of WMVDs. If d is trivial, i.e., $Y \subset X$, then it follows from rules W1 and W2. Otherwise, we note that WMVDs are full TDs. Hence, by Lemma 1, D logically implies d as a TD $(t_1,t_2,t_3)/t_4$ if and only if chasing t_1,t_2,t_3 under D causes the generation of t_4. By Lemma 3, we know that a successful chase sequence could produce t_4 in one step. By Lemma 2, $X \text{-w->} Y$ can be inferred from D using the rules W2, W3, W4.

4. <u>Applications of Weak MVDs</u>

4.1 Nesting.

The concept of nesting is a useful notion in the study of non-first-normal-form relations. Nesting on a single attribute was first studied in [JS] and was generalized in [SP] and in [T]. While Jaeschke and Scheck looked at the problem of having two NEST operations commute, they did not consider this problem for three or more NESTs [JS]. We give a solution below.

<u>Definition 2</u>.

Let U be a set of objects and r a relation over U. In this paper an object will either be an ordinary attribute or a nested set of attributes.

Let $\emptyset \neq X \subseteq U$ and $Y = U-X$

$$NEST_X\ (U,r)=(U_X,r_X)$$

where U_X is a set of objects, equal to $(U-X)\ \bar{X}$.
Thus the set of objects X is replaced by a new single object $\bar{X}$.

$r_X = \{t \mid$ there exists a tuple $u \in \Pi_Y(r)$ such that $t[Y]=u$ and $t[X]=\{v[X] \mid v \in r$ and $v[Y]=u\}\}$.

Example

Consider the relation

$(U=\{EMP\#, CHILDNAME, CHILDAGE\}, r)$ in Figure 1.

EMP#	CHILDNAME	CHILDAGE
2510	RALPH	10
4510	SAM	15
3614	GAIL	3
2510	BETH	25
4910	ANN	17

Figure 1

EMP#	CHILDINFO
2510	{(RALPH,10), (BETH,25)}
4910	{(SAM, 15), (ANN, 17)}
3614	{(GAIL,3)}

Figure 2

$Nest_{\{CHILDNAME,\ CHILDAGE\}}(U,r)=$

$(\{EMP\#, CHILDINFO\}, r')$ shown in figure 2.

Weak multivalued dependencies were introduced by Jaeschke and Scheck in order to characterize when two NEST operations on different single attribute would commute [JS]. Thomas extended this result to nesting on arbitrary objects [T]. Their findings are summarized in the following lemma.

Lemma 4 [JS,T].

Let U be a set of objects and r a relation over U. Let X,Y,Z be subsets of U such that Z=U-XY.

X -w-> Y holds in r

if and only if

$NEST_Y(NEST_Z(r))=NEST_Z(NEST_Y(r))$

We can generalize this lemma to deal with nesting on more than two sets of objects. We first prove the following lemma.

Lemma 5.

Let U be a set of objects (some of which may be nested sets of attributes) and r a relation over U. Let $Y_1, \ldots Y_n$ be pairwise disjoint subsets of U $(n \geq 3)$.

If for any $1 \leq i < j \leq n$, $(U-Y_iY_j)$ -w-> Y_i holds in (U,r) then for any $1 \leq i < j \leq n$, $i \neq k$ and $j \neq k$ $(U_k-Y_iY_j)$ -w-> Y_i holds in (U_k, r_k) where

$$U_k = XY_1 \ldots Y_{k-1}\overline{Y}_kY_{k+1} \ldots Y_n$$
$$r_k = NEST_{Y_k}(r)$$

Proof. Assume the given WMVDs hold in r and define $X=U-Y_1Y_2 \ldots Y_n$. We have to prove that for i,j,k distinct, $(U_k-Y_iY_j)$ -w-> Y_i holds in r_k. This is equivalent to proving that the following template dependency holds in r_k

$$\begin{array}{l}
X \quad Y_1 \ldots Y_i \ldots Y_{k-1} \quad \overline{Y}_k \; Y_{k+1} \ldots Y_j \ldots Y_n \\
\hline
x \quad y_1 \ldots y_i \ldots y_{k-1} \quad S_k \; y_{k+1} \ldots y_j \ldots y_n \\
x \quad y_1 \ldots y_i \ldots y_{k-1} \quad S_k \; y_{k+1} \ldots y_j' \ldots y_n \\
x \quad y_1 \ldots y_i' \ldots y_{k-1} \quad S_k \; y_{k+1} \ldots y_j \ldots y_n \\
\hline
x \quad y_1 \ldots y_i' \ldots y_{k-1} \quad S_k \; y_{k+1} \ldots y_j' \ldots y_n
\end{array}$$

Since $(U-Y_iY_j)$ -w-> Y_i holds in r we know that for each $y_k \in S_k$, r contains the tuple $x\, y_1 \ldots y_i' \ldots y_{k-1}\, y_k\, y_{k+1} \ldots y_j' \ldots y_n$. Hence, by the definition of NEST r_k contains the supertuple $x\, y_1 \ldots y_i' \ldots y_{k-1}\, S\, y_{k+1} \ldots y_j' \ldots y_n$, where $S \subseteq S_k$. We have to prove that $S=S_k$ (note that $S_k \neq \emptyset$ by the definition of NEST).

Suppose $y \in S$ and let $y_k \in S_k$. Then the following tuples belong to r

$$\begin{array}{l}
x \quad y_1 \ldots y_i' \ldots y_{k-1} \quad y \quad y_{k+1} \ldots y_j' \ldots y_n \\
x \quad y_1 \ldots y_i' \ldots y_{k-1} \quad y_k \quad y_{k+1} \ldots y_j' \ldots y_n \\
x \quad y_1 \ldots y_i' \ldots y_{k-1} \quad y_k \quad y_{k+1} \ldots y_j \ldots y_n
\end{array}$$

Since $(U-Y_kY_j)$ -w-> Y_k holds in r, r also contains the tuple $x\, y_1 \ldots y_i' \ldots y_{k-1}\, y\, y_{k+1} \ldots y_j \ldots y_n$. Hence, by the definition of NEST, $y \in S_k$.

Theorem 3. Let U be a set of attributes and r a relation over U. Let $Y_1, Y_2, \ldots, Y_n$ be nonempty pairwise disjoint subsets of U. Let N_i abbreviate $NEST_{Y_i}$. The WMVDs $(U-Y_iY_j)$ -w-> Y_i all hold in r

for $1 \leq i < j \leq n$ if and only if for any permutation $i_1, i_2, \ldots i_n$ of the integers $1,2,\ldots n$,

$$N_1(N_2(\ldots(N_n(r))\ldots))=$$

$$N_{i_1}(N_{i_2}(\ldots(N_{i_n}(r)\ldots)).$$

Proof. For the "if" part consider permutations which are identical except the first two nests are on Y_i and Y_j in one permutation, Y_j and Y_i in the other. By unnesting [T] we obtain

$N_i(N_j(r)) = N_j(N_i(r))$.

The result then comes from Lemma 4. Since any permutation may be generated by a sequence of adjacent pair interchanges, to verify the "only if" part it is sufficient to show that for any permutation $i_1, i_2, \ldots i_n$ of $1,2,\ldots,n$ and any $1 \leq k < n$.

$$N_{i_1}(N_{i_2}(\ldots N_{i_k}(N_{i_{k+1}}(\ldots(N_{i_n}(r))\ldots))\ldots))=$$

$$N_{i_1}(N_{i_2}(\ldots N_{i_{k+1}}(N_{i_k}(\ldots(N_{i_n}(r))\ldots))\ldots))$$

From Lemma 5, it follows that the WMVD $(U^*-Y_{i_k}Y_{i_{k+1}}) \text{-w->} Y_{i_k}$ holds in

$$N_{i_{k+2}}(N_{i_{k+3}}(\ldots(N_{i_n}(U,r))..))=(U^*,r^*)$$

where U^* is the set of objects obtained after nesting on $Y_{i_n}, Y_{i_{n-1}}, \ldots, Y_{i_{k+2}}$ and r^* is the respective instance.

Then by Lemma 4,

$N_{i_k}(N_{i_{k+1}}(U^*,r^*)=N_{i_{k+1}}(N_{i_k}(U^*,r^*)$. The derived equality follows immediately.

4.2 Decompositions.

Horizontal decompositions were proposed in [AD] and [F] as a way of breaking a relation into smaller relations where more desirable properties might hold locally even though they did not hold globally. More recent work on two-way horizontal decompositions is given in [KY] and [PD]. We review and characterize certain two-way horizontal decompositions in Section 5 below. We discuss here a horizontal decomposition obtained by nesting.

Let $Y_1, Y_2, \ldots, Y_n$ be disjoint sets of attributes and $X=U-Y_1Y_2\ldots Y_n$. If one performs nest operations on a relation r on all of the Y_i's, one will obtain the following structure:

	X	Y_1	Y_2	...	Y_n
t_1:	x_1	S_{11}	S_{12}	...	S_{1n}
t_2:	x_2	S_{21}	S_{22}	...	S_{2n}
			.		
			.		
			.		
t_k:	x_k	S_{k1}	S_{k2}	...	S_{kn}

where S_{ij} is a set of tuples over Y_j. In general, the number k of "super-tuples" and the composition of the S_{ij}'s will depend upon the order of nesting; indeed, many of the sets S_{ij} may turn out to be singletons. However, no information is lost since the structure could be unnested back to 1NF (cf. [JS,B,T]). Clearly each S_{ij}, whatever its size, associates a nonempty set of Y_j-values with the X-value x_i. Furthermore, each super-tuple t_i can represent those tuples in the original relation r which can be obtained by unnesting t_i. Thus r can be partitioned horizontally into subrelations $r_1, r_2, \ldots, r_k$. In general, this decomposition may be as complex as r itself.

When the conditions of Theorem 3 hold, however, the above scheme becomes more useful. First, the sets S_{ij} become unique (i.e., independent of the ordering of nesting). Second, the number k tends to become smaller. Third, each subrelation r_i can be vertically decomposed further via the scheme $(XY_1, XY_2, \ldots, XY_n)$ in lossless fashion. This local join dependency (JD) arises from the nesting process since the choice of the r_i depends upon the relation (instance) r. Although the same local JD holds for each of the r_i's, information must be kept as to which r_i produced which subrelations. The same JD clearly does not hold for all of r since the x_i's are not necessarily distinct.

4.3 Query Processing.

It is well known that a subclass of the relational queries, the so-called tree queries, can be processed efficiently. The problem of transforming cyclic queries into tree queries has been studied in [GS,KY]. Goodman and Shmueli use techniques such as attribute deletion and addition. Kambayashi and Yoshikawa employ FDs, MVDs and their version of degenerate MVDs to establish the transformation. WMVDs can be added to the above dependencies in a natural way.

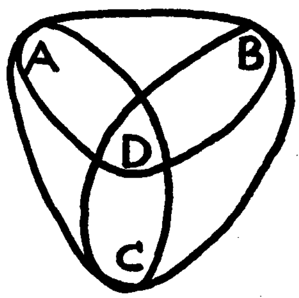
The class of queries Kambayashi and Yoshikawa consider are the so-called natural queries [KY]. These queries (q) have the form $\Pi_T(R_1,r_1)|X|\ldots|X|(R_n,r_n)$, where T, $R_1, \ldots, R_n$ are subsets of U, and for $1 \leq i \leq n$ r_i is a relation over R_i. T is called the target. We assume the r_i's reside at different sites of a distributed database system. The query-hypergraph $\mathcal{H}_q=(\mathcal{E}_q,\mathcal{N}_q)$ associated with q is defined as follows

$\mathcal{N}_q = U$ the set of all attributes

$\mathcal{E}_q = \{R_i \mid i=1,\ldots,n\}$ the set of relation schemas in q

A natural query q is a tree query if ${}_q$ is acyclic (cf. [BFMY]) otherwise it is called cyclic. The main objective in [KY] was to illustrate how dependencies could be used to transform a cyclic query into a tree query. As an example, consider the following cyclic query

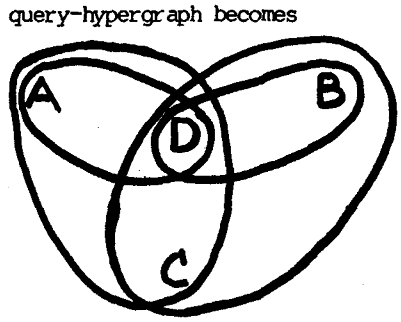
$q=((R_1=ABD,r_1)|X|(R_2=BCD,r_2)|X|(R_3=ACD,r_3))$ and its query-hypergraph



Suppose the MVD D ->-> A holds in (R_1, r_1). We can then decompose (R_1, r_1) into $(R_{11} = AD, r_{11})$ and $(R_{12} = BD, r_{12})$. Then q becomes

$((R_{11},r_{11})|X|(R_{12},r_{12})|X|(R_2,r_2)|X|(R_3,r_3))$

and the query-hypergraph becomes



This hypergraph is acyclic so q can be evaluated efficiently. Suppose however that instead of the MVD D ->-> A, the WMVD D -w-> A holds in (R_1,r_1). Straight vertical decomposition is not possible. However we can employ Theorem 1.4. Since D -w-> A holds in r_1, we can decompose it into a set of subrelations $s_{11},s_{12},\ldots s_{1n}$, where D ->-> A holds in each s_{1i}. We could then perform the same analysis as above on each subrelation (R_1,s_{1i}) to produce

$q=\bigcup_{i=1}^{n}((R_{11}, s_{1i1})|X|(R_{12},s_{1i2})|X|(R_2,r_2)|X|(R_3,r_3))$.

Thus q has been transformed into a union of tree queries.

4.4 Addition of Attributes.

Sciore has considered the problem of improving database design by adding attributes [S]. Adding an attribute when a nested scheme as in 4.1, 4.2 above is obtained may be useful. This attribute A would essentially be a "super-tuple i.d." In this case, the MVD XA ->-> Y would hold globally and the single vertical decomposition (XAY,XAZ) would be possible. This attribute can have a natural meaning in some cases: e.g., in a student, course, teacher database, A could have the meaning "section number".

5. <u>Degenerate Multivalued Dependencies</u>.

We wish to compare weak MVDs with "degenerate" MVDs, which appear in [AD] and [SDPF]. A different form of "degenerate" MVD was recently given in [KY] and is also analyzed.

<u>Definition 3</u>. A <u>degenerate</u> <u>MVD</u> (DMVD), written X -d-> Y, is satisfied by a relation r if and only if for any pair t, t' of tuples of r, if t[X] = t'[X], then either t[Y] = t'[Y] or t[Z] = t'[Z], where Z = U-XY as before.

<u>Remark 3</u>. [AD,JS]. The four notions, X --> Y, X -d-> Y, X ->-> Y, X -w-> Y, become successively and strictly weaker, i.e., easier to satisfy, as one progresses from FDs to WMVDs.

The "degenerate" MVD given in [KY] is defined as follows:

X -ky-> Y holds in r if and only if r can be partitioned horizontally into two relations, r_1 and r_2 such that the FD X --> Y holds in r_1 and the FD X --> Z holds in r_2, where Z = U-XY.

<u>Remark 4</u>. If r satisfies X -d-> Y, then r satisfies X -ky-> Y. The converse does not hold. Furthermore, the notion given in [KY] is incomparable with either ordinary or weak MVDs.

<u>Proof</u>. If X -d-> Y, then all tuples of r with a common X-value must satisfy either X --> Y or X --> Z. The relation consisting of the first four rows of r_{ex} in Section 2 satisfies X ->-> Y but not X -ky-> Y. The relation consisting of the first three rows of r_{ex} satisfies X -ky-> Y but not X -w-> Y.

In terms of the approach of Beeri and Vardi [BV], DMVDs are equality-generating dependencies (as are FDs), while WMVDs are tuple-generating dependencies (as are MVDs). The [KY] "dependency" is more complex and is really a global property of relations. Its characterization is negative in character as we see next.

<u>Theorem 4</u>. Let U, X, Y, Z and r be as in Theorem 1. Then X -ky-> Y fails to hold in r if and only if r contains a subrelation r' of one of the following forms:

Type I. r' consists of three tuples t_1,t_2,t_3 with
- a. $t_1[X] = t_2[X] = t_3[X]$;
- b. $t_1[Y]$, $t_2[Y]$, $t_3[Y]$ are all distinct;
- c. $t_1[Z]$, $t_2[Z]$, $t_3[Z]$ are all distinct.

Type II. r' consists of four distinct tuples t_1,t_2,t_3,t_4 with
- a. $t_1[X] = t_2[X] = t_3[X] = t_4[X]$;
- b. $t_1[Y] = t_2[Y] \neq t_3[Y] = t_4[Y]$.

Type III. r' consists of four distinct tuples t_1,t_2,t_3,t_4 with
- a. $t_1[X] = t_2[X] = t_3[X] = t_4[X]$;
- b. $t_1[Z] = t_2[Z] \neq t_3[Z] = t_4[Z]$.

<u>Proof</u>. The "if" part is straightforward. The "only if" part proceeds by observing that if X -ky-> Y is not satisfied by r, there must be a particular X-value for which it fails, i.e., it is not satisfied by $s' = \sigma_{X=x}(r)$ for some x. There must be a maximal subrelation s of s' for which the property still holds via subrelations s_1, s_2. The conclusion follows by analysis of s_1, s_2, and any tuple in s'-s where s_1 satisfies X —> Y and s_2 satisfies X —> Z.

Let t be any tuple of s'-s. Then we have the following formats for s_1, s_2 and t:

	X	Y	Z
	x	y_0	z_1
	x	y_0	z_2
s_1:		.	
		.	
		.	
	x	y_0	z_m
	x	y_1	z_0
	x	y_2	z_0
s_2:		.	
		.	
		.	
	x	y_n	z_0
t:	x	y	z

Clearly $z_1,z_2,\ldots,z_m$ are distinct, as are $y_1,y_2,\ldots,y_n$. We leave it to the reader to complete the proof by applying the fact that $s_1 \cup s_2 \cup \{t\}$ violates X -ky-> Y to the four cases: $m=n=1$; $m>1=n$; $m=1<n$; $m>1$ and $n>1$.

6. <u>Mixed Systems involving WMVDs</u>.

In the previous section we showed that a system of WMVDs is not very powerful in the sense that "essentially different" WMVDs cannot be generated indirectly from a given set of WMVDs. This situation changes, however, when both WMVDs and MVDs are present. We present two mixed inference rules, one of which suggests a form of transitivity. We conjecture that the inference rules for WMVDs (W1 through W4), for MVDs as in [B] and the mixed rules below form a complete axiomatization for systems of MVDs and WMVDs, but we have not yet proven this.

WM1. Generalization. If X ->-> Y, then X -w-> Y.

WM2. Weak transitivity. Let W,X,Y,Z be disjoint subsets of U such that U = WXYZ. If XW -w-> Y, XY -w-> Z, XZ ->-> Y, then X -w-> Z.

WM2 is a somewhat unnatural axiom, but is strictly stronger than the more natural mixed-transitivity principle: X ->-> Y and Y -w-> Z implies X -w-> Z. In WM2, it is necessary that the third condition be an ordinary MVD. On the other hand, if all three conditions of WM2 are MVDs, the consequent is still only a WMVD. Note that it is false that X -w-> Y and Y ->-> Z implies X -w-> Z. We also note that WMVDs do not interact in a special way with FDs. This follows from a result in [SU2].

Two open problems are:

1. Prove completeness of the system for WMVDs and MVDs (or find the right axiom system, if necessary).
2. See if there is an analgous concept of a "weak join dependency" (WJD). Compare this to the conditions in Theorem 2. See if the notion of acyclicity has an analogue for WJDs.

Bibliography

[A] W.W. Armstrong, Dependency Structures of Database Relationships. Information Processing 74 (North-Holland, Amsterdam, 1974), 580-583.

[AD] W.W. Armstrong, C. Delobel, Decompositions and Functional Dependencies in Relations. ACM Trans. on Database Systems Vol. 5, No. 4 (Dec. 1980), 404-430.

[B] C. Beeri, On the Membership Problem for Functional and Multivalued Dependencies in Relational Databases. ACM Trans. on Database Systems, Vol. 5, No. 3 (Sept. 1980), 241-259.

[BFMY] C. Beeri, R. Fagin, D. Maier, M. Yannakakis, On the Desirability of Acyclic Data base Schemes. J. ACM, Vol. 28, No. 3 (July 1983), 479-513.

[BV] C. Beeri, M.Y. Vardi, Formal Systems for Tuple and Equality Generating Dependencies. Tech. Report,to appear, Hebrew University of Jerusalem, April, 1981.

[F] R. Fagin, Normal Forms and Relational Database Operators. Proc. ACM-SIGMOD 1979 Int'l. Conf. Management of Data, Boston, Mass., 153-160.

[FMU] R. Fagin, A.O. Mendelzon, J.D. Ullman, A Simplified Universal Relation Assumption and Its Properties. ACM Trans. on Database Systems, Vol. 7, No. 3 (Sept. 1982), 343-360.

[FMUY] R. Fagin, D. Maier, J.D. Ullman, M. Yannakakis, Tools for Template Dependencies. SIAM J. Computing, Vol. 12, No. 1 (Feb. 1983), 36-59.

[FT] P.C. Fischer, S. Thomas, Operators for Non-First-Normal Form Relations. Proc. 1983 Computer Software and Applications Conf., 464-475.

[GS] N. Goodman, O. Shmueli, Transforming Cyclic Schemas into Trees. Proc. First ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, 1982, 49-54.

[JS] G. Jaeschke, H.J. Scheck, Remarks on the Algebra of Non-First-Normal-Form Relations. Proc. First ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, 1982, 124-138.

[KY] Y. Kambayashi, M. Yoshikawa, Query Processing Utilizing Dependencies and Horizontal Decomposition. Proc. ACM SIGMOD 1983 Int'l. Conf. on Management of Data, 55-67.

[PD] J. Paredaens, P. DeBra, On Horizontal Decomposition, Proc. XP2 Workshop on Relational Database Theory, Pennsylvania State University, 1981.

[S] E. Sciore, Improving Database Schemes by Adding Attributes, Proc. Second ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, 1983, 363-372.

[SDPF] Y. Sagiv, C. Delobel, D.S. Parker, R. Fagin, An Equivalence Between Relational Database Dependencies and a Fragment of Propositional Logic. J. ACM, Vol. 28, No. 3, (July 1981), 435-453.

[SP] H.J. Scheck, P. Pistor, Data Structures for an Integrated Data Base Management and Information Retrieval Systems. Proc. on Very Large Data Bases, 1982, 197-207.

[SU1] F. Sadri, J.D. Ullman, Template Dependencies: A Large Class of Dependencies in Relational Databases and its Complete Axiomatization. J. ACM, Vol. 28, No. 3, (July 1981), 435-453.

[SU2] F. Sadri, J.D. Ullman, The Interaction Between Functional Dependencies and Template Dependencies. Proc. ACM-SIGMOD 1980 Int'l. Conf. on Management of Data, 45-51.

[T] S. Thomas, A Non-First-Normal-Form Relational Database Model. Doctoral thesis, Vanderbilt University, 1983.

[U] J.D. Ullman, Principles of Database Systems, Computer Science Press, Rockville, MD, 1980.