Annales Univ. Sci. Budapest., Sect. Comp. 19 (2000) 119-132

SOME COMPUTATIONAL PROBLEMS RELATED TO BOYCE-CODD NORMAL FORM

J. Demetrovics and Vu Duc Thi
(Budapest, Hungary)

Abstract. In the relational database theory the most desirable normal form is the Boyce-Codd normal form (BCNF). This paper investigates some computational problems concerning BCNF relation schemes and BCNF relations. We give an effective algorithm finding a BCNF relation r such that r represents a given BCNF relation scheme s (i. e. $K_r = K_s$, where K_r and K_s are sets of all minimal keys of r and s). This paper also gives an effective algorithm which from a given BCNF relation finds a BCNF relation scheme such that $K_r = K_s$. Based on these algorithms we prove that the time complexity of the problem, that finds a BCNF relation r representing a given BCNF relation scheme s, is exponential in the size of s and conversely, the complexity of finding a BCNF relation scheme s from a given BCNF relation r such that r represents s also is exponential in the number of attributes.

We give a new characterization of the relations and the relation schemes that are uniquely determined by their minimal keys. It is known that these relations and relation schemes are in the BCNF class. From this characterization we give a polynomial time algorithm deciding whether an arbitrary relation is uniquely determined by its set of all minimal keys. In the rest of this paper some new bounds of the size of minimal Armstrong relations for BCNF relation schemes are given. We show that given a Sperner system K and BCNF relation scheme s a set of minimal keys of which is K, the number of antikeys (maximal nonkeys) of K is polynomial in the number of attributes iff so is the size of minimal Armstrong relation of s.

Research supported by Hungarian Foundation for Scientific Research Grant 2575.

1. Introduction

In the relational datamodel one of the important concepts is the functional dependency (FD). Several types of families of FDs which satisfy some conditions are known under the name normal forms (NFs). The Boyce-Codd normal form has been investigated in a lot of papers. It is shown [4] that every set of attributes with an associated set of FDs has a decomposition into third NF which has the loss-less-join property and preserves FDs. However, for BCNF this does not always exist. The key is an interesting concept in the relational datamodel. In this paper we present some computational problems related to connections between sets of minimal keys, relation schemes and relations in BCNF class.

Let us give some necessary definitions and results that are used in the next section.

Definition 1.1. Let $R = \{a_1, \ldots, a_n\}$ be a nonempty finite set of attributes, $r = \{h_1, \ldots, h_m\}$ be a relation over R, and $A, B \subseteq R$. Then we say that B functionally depends on A in r (denoted $A \xrightarrow[r]{f} B$) iff

$$(\forall h_i, h_j \in r)((\forall a \in A)(h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)(h_i(b) = h_j(b))).$$

Let $F_r = \{(A, B) : A, B \subseteq R, A \xrightarrow[r]{f} B\}$. F_r is called the full family of functional dependencies of r, where we write (A, B) or $A \rightarrow B$ for $A \xrightarrow[r]{f} B$ when r, f are clear from the context.

Definition 1.2. A functional dependency over R is a statement of the form $A \rightarrow B$, where $A, B \subseteq R$. The FD $A \rightarrow B$ holds in a relation r if $A \xrightarrow[r]{f} B$. We also say that r satisfies the FD $A \rightarrow B$. Clearly, F_r is a set of all FDs that hold in r.

Definition 1.3. Let R be a nonempty finite set, and denote $P(R)$ its power set. Let $y \subseteq P(R) \times P(R)$. We say that y is an f-family over R iff for all $A, B, C, D \subseteq R$

(1) $(A, A) \in y$,

(2) $(A, B) \in y,\ (B, C) \in y \Rightarrow (A, C) \in y$,

(3) $(A, B) \in y,\ A \subseteq C,\ D \subseteq B \Rightarrow (C, D) \in y$,

(4) $(A, B) \in y,\ (C, D) \in y \Rightarrow (A \cup C,\ B \cup D) \in y$.

Clearly, F_r is an f-family over R.

It is known [1] that if y is an arbitrary f-family, then there is a relation r over R such that $F_r = y$.

Definition 1.4. A relation scheme s is a pair $\langle R, F\rangle$, where R is a set of attributes, and F is a set of FDs over R. Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 1.3. Denote $A^+ = \{a : A \to \{a\} \in \in F^+\}$. A^+ is called the closure of A over s. It is clear that $A \to B \in F^+$ iff $B \subseteq A^+$.

Clearly, if $s = \langle R, F\rangle$ is a relation scheme, then there is a relation r over R such that $F_r = F^+$ (see [1]). Such a relation is called an Armstrong relation of s. It is obvious that all FDs of s hold in r.

Definition 1.5. Let r be a relation, $s = \langle R, F\rangle$ be a relation scheme, y be an f-family over R and $A \subseteq R$. Then A is a key of r (a key of s, a key of y) if $A \xrightarrow[r]{f} R$ $(A \to R \in F^+, (A, R) \in y)$. A is a minimal key of $r(s, y)$ if A is a key of $r(s, y)$ and any proper subset of A is not a key of $r(s, y)$. Denote K_r (K_s, K_y) the set of all minimal keys of $r(s, y)$. Clearly, K_r, K_s, K_y are Sperner systems over R (i.e. $A, B \in K_r$ implies $A \not\subseteq B$).

Definition 1.6. Let K be a Sperner system over R. We define the set of antikeys of K, denoted by K^{-1}, as follows

$$K^{-1} = \{A \subset R : (B \in K) \Rightarrow (B \not\subseteq A) \text{ and } (A \subset C) \Rightarrow (\exists B \in K)(B \subseteq C)\}.$$

It is easy to see that K^{-1} is also a Sperner system over R.

It is known [6] that if K is an arbitrary Sperner system over R then there is a relation scheme s such that $K_s = K$.

In this paper we always assume that if a Sperner system plays the role of the set of minimal keys (antikeys), then this Sperner system is not empty (does not contain R). We consider the comparison of two attributes as an elementary step of algorithms. Thus, if we assume that subsets of R are represented as sorted lists of attributes, then a Boolean operation on two subsets of R requires at most $|R|$ elementary steps.

Definition 1.7. Let $I \subseteq P(R)$, $R \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(R)$. Denote $M^+ = \{\cap M' \ : \ M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $R \in M^+$ but not in M, since it is the intersection of the empty collection of sets.

Denote $N = \{A \in I : \ A \neq \cap\{A' \in I : \ A \subset A'\}\}$.

In [6] it is proved that N is the unique minimal generator of I. Thus, for any generator N' of I we obtain $N \subseteq N'$.

Definition 1.8. Let r be a relation over R, and E_r the equality set of r, i.e. $E_r = \{E_{ij} \ : \ 1 \leq i < j \leq |r|\}$. where $E_{ij} = \{a \in R \ : \ h_i(a) = h_j(a)\}$.

Let $T_r = \{A \in P(R) : \exists E_{ij} = A, \nexists E_{pq} \quad A \subset E_{pq}\}$. Then T_r is called the maximal equality system of r.

Definition 1.9. Let r be a relation, and K a Sperner system over R. We say that r represents K iff $K_r = K$.

The following theorem is known ([8]).

Theorem 1.10. *Let K be a non-empty Sperner system and r a relation over R. Then r represents K iff $K^{-1} = T_r$, where T_r is the maximal equality system of r.*

Let $s = \langle R, F \rangle$ be a relation scheme over R, K_s is a set of all minimal keys of s. Denote K_s^{-1} the set of all antikeys of s. From Theorem 1.10 we obtain the following corollary.

Corollary 1.11. *Let $s = \langle R, F \rangle$ be a relation scheme and r a relation over R. We say that r represents s if $K_r = K_s$. Then r represents s iff $K_s^{-1} = T_r$, where T_r is the maximal equality system of r.*

In [7] we proved the following theorem.

Theorem 1.12. *Let $r = \{h_1, \ldots, h_m\}$ be a relation, and F an f-family over R. Then $F_r = F$ iff for every $A \in P(R)$*

$$H_F(A) = \begin{cases} \bigcap\limits_{A \subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij} \in E_r : A \subseteq E_{ij}, \\ R & \text{otherwise,} \end{cases}$$

where $H_F(A) = \{a \in R : (A, \{a\}) \in F\}$ and E_r is the equality set of r.

We say that a relation scheme $s = \langle R, F \rangle$ (a relation r) is in BCNF if $\forall A \subseteq R$ either $A^+ = A$ or $A^+ = R$ ($H_{F_r}(A) = A$ or $H_{F_r}(A) = R$).

2. Results

All relations and relation schemes investigated in this section are in BCNF. First we construct two combinatorial algorithms concerning minimal keys of relations and relation schemes. We estimate these algorithms. After that we present two problems the worst-case time complexity of which are exponential.

Let $s = \langle R, F \rangle$ be a relation scheme over R. From s we construct $Z(s) = \{X^+ : X \subseteq R\}$, and compute the minimal generator N_s of $Z(s)$. We put

$$T_s = \{A \in N_s : \nexists B \in N_s : A \subset B\}.$$

It is known [1] that for a given relation scheme s there is a relation r such that r is an Armstrong relation of s. On the other hand, by Corollary 1.11 and Theorem 1.12, the following proposition is clear.

Proposition 2.1. *Let $s = \langle R, F\rangle$ be a relation scheme over R. Then*

$$K_s^{-1} = T_s.$$

Definition 2.2. Let $s = \langle R, F\rangle$ be a relation scheme. We say that s is a k-relation scheme over R if $F = \{K_1 \to R, \ldots, K_m \to R\}$, where $\{K_1, \ldots, K_m\}$ is a Sperner system over R. It is easy to see that $K_s = \{K_1, \ldots, K_m\}$.

Clearly, if $s = \langle R, F\rangle$ is in BCNF then using the algorithm for finding a minimal cover in polynomial time we can construct a k-relation scheme $s' = \langle R, F'\rangle$ such that $F^+ = F'^+$, see [10]. Conversely, it can be seen that an arbitrary k-relation scheme is in BCNF. Consequently, we can consider a relation scheme in BCNF as a k-relation scheme.

Remark 2.3. It is known [10] that $s = \langle R, F\rangle$ is in BCNF iff its minimum cover is a k-relation scheme. Consequently, the BCNF property of s is polynomially recognizable.

Let r be a relation over R. From r we compute E_r. We construct the maximal equality system T_r of r. By Theorem 1.10 we obtain $T_r = K_r^{-1}$. Denote elements of T_r by $A_1, \ldots, A_t$.

Set $M_r = \{A_i - a \ : \ a \in R, \ i = 1, \ldots, t\}$. Denote elements of M_r by $B_1, \ldots, B_s$. We construct the relation $r' = \{h_0, h_1, \ldots, h_s\}$ as follows:

for all $a \in R$, $h_0(a) = 0$, for each $i = 1, \ldots, s$ $h_i(a) = 0$ if $a \in B_i$, in the converse case we set $h_i(a) = i$.

By [10] r' is in BCNF and $K_r = K_{r'}$ (1). It is easy to see that M_r and r' are constructed in polynomial time in the size of r.

Set $H_{F_r}(A) = \{a \in R \ : \ (A, \{a\}) \in F_r\}$, $Z_{F_r} = \{A \quad H_{F_r}(A) = A\}$. Denote by N_{F_r} the minimal generator of Z_{F_r}.

Based on definition of BCNF relation and from (1) we can see that a relation r is in BCNF iff $N_{F_r} = N_{F_{r'}}$. Because for an arbitrary relation r N_{F_r} is computed in polynomial time, the BCNF property of r can be tested in polynomial time.

We give the following algorithm that from a given relation scheme s constructs a relation r such that r represents s. It is known [15] that there is an algorithm that finds a set of all antikeys from a given Sperner system.

Algorithm 2.4. (Finding a set of antikeys)

Input: Let $K = \{B_1, \ldots, B_m\}$ be a Sperner systems over R

Output: K^{-1}

Step 1: We set $K_1 = \{R - \{a\} : a \in B_1\}$. It is obvious that $K_1 = \{B_1\}^{-1}$

Step $q+1$: $(q < m)$ We assume that $K_q = F_q \cup \{X_1, \ldots, X_{t_q}\}$, where $X_1, \ldots, X_{t_q}$ containing B_{q+1} and $F_q = \{A \in K_q : B_{q+1} \not\subseteq A\}$. For all i $(i = 1, \ldots, t_q)$ we construct the antikeys of $\{B_{q+1}\}$ on X_i in an analogous way as K_1. Denote them by $A_1^i, \ldots, A_{r_i}^i$ $(i = 1, \ldots, t_q)$. Let

$$K_{q+1} = F_q \cup \{A_p^i : A \in F_q \Rightarrow A_p^i \not\subseteq A,\ 1 \le i \le t_q,\ 1 \le p \le r_i\}.$$

We set $K^{-1} = K_m$.

Theorem 2.5. *For every q $(1 \le q \le m)$, $K_q = \{B_1, \ldots, B_q\}^{-1}$, i.e. $K_m = K^{-1}$.*

It can be seen that K and K^{-1} are uniquely determined by one another and the determination of K^{-1} based on our algorithm does not depend on the order of $B_1, \ldots, B_m$. Denote $K_q = F_q \cup \{X_1, \ldots, X_{t_q}\}$ and let l_q $(1 \le q \le m-1)$ be the number of elements of K_q.

Proposition 2.6. [15] *The worst-case time complexity of Algorithm 2.4 is*

$$O\left(|R|^2 \sum_{q=1}^{m-1} t_q u_q\right),$$

where

$$u_q = \begin{cases} l_q - t_q & \text{if} \quad l_q > t_q, \\ 1 & \text{if} \quad l_q = t_q. \end{cases}$$

Clearly, in each step of our algorithm K_q is a Sperner system. In the cases for which $l_q \le l_m$ $(q = 1, \ldots, m-1)$ it is easy to see that the time complexity of our algorithm is not greater than $O(|R|^2|K||K^{-1}|^2)$. Thus, in these cases Algorithm 2.4 finds K^{-1} in polynomial time in $|R|, |K|$ and $|K^{-1}|$. It can be seen that if the number of elements of K is small then Algorithm 2.4 is very effective. It only requires polynomial time in $|R|$.

By Algorithm 2.4 we construct an algorithm that finds a relation such that this relation represents a given relation scheme. By Remark 2.3 it is simple that we can consider an arbitrary relation scheme in BCNF as a k-relation scheme.

Algorithm 2.7.

Input: $s = \langle R, F = \{K_1 \to R, \ldots, K_m \to R\}\rangle$ be a k-relation scheme

Output: a BCNF relation r such that $K_r = K_s$

Step 1: From $K = \{K_1, \ldots, K_m\}$ we contruct $K^{-1} = \{B_1, \ldots, B_t\}$ by Algorithm 2.4

Step 2: Set $M = \{B_i - a \quad a \in R,\ i = 1, \ldots, t\}$

Step 3: Denote elements of M by $A_1, \ldots, A_l$, construct a relation $r = \{h_0, h_1, \ldots, h_l\}$ as follows: for all $a \in R$: $h_0(a) = 0$. For $i = 1, \ldots, l$ we set $h_i(a) = 0$ if $a \in A_i$, in the converse case $h_i(a) = i$.

By Remark 2.3, Corollary 1.11 we obtain $K_r = K_s$, and r is a BCNF relation.

Clearly, set M and relation r are constructed in polynomial time in the size of K^{-1}. Consequently, the time complexity of this algorithm is $O\left(|R|^3 \sum_{q=1}^{m-1} t_q u_q\right)$, for meanings of t_q, u_q see Proposition 2.6. In many cases this algorithm requires polynomial time in the size of s (see Proposition 2.6).

Now we construct an algorithm which finds a BCNF relation scheme such that a given BCNF relation represents this relation scheme. First we give the following algorithm.

Algorithm 2.8. [8] (Finding a minimal key from a set of antikeys)

Input: Let K be a Sperner system, H a Sperner system, and $C = \{b_1, \ldots, b_m\} \subseteq R$ such that $H^{-1} = K$ and $\exists B \in K$: $B \subseteq C$.

Output: $D \in H$

Step 1: Set $T(0) = C$

Step $i + 1$: Set $T = T(i) - b_{i+1}$.

$$T(i+1) = \begin{cases} T & \text{if } \forall B \in K \ : \ T \not\subseteq B, \\ T(i) & \text{otherwise.} \end{cases}$$

We set $D = T(m)$.

Lemma 2.9. *[8] If K is a set of antikeys, then $T(m) \in H$.*

Lemma 2.10. *[8] Let H be a Sperner system over R, and $H^{-1} = \{B_1, \ldots, B_m\}$ be a set of antikeys of $H, T \subseteq H$. Then $T \subset H$, $T \neq \emptyset$ if and only if there is a $B \subseteq R$ such that $B \in T^{-1}$, $B \not\subseteq B_i$ $(\forall i \ : \ 1 \leq i \leq m)$.*

Based on Lemma 2.10, Algorithm 2.8 we have the following algorithm.

Algorithm 2.11. (Finding a set of minimal keys from a set of antikeys)

Input: Let $K = \{B_1, \ldots, B_k\}$ be a Sperner system over R.

Output: H such that $H^{-1} = K$

Step 1: By Algorithm 2.8 we compute an A_1, set $K(1) = A_1$

Step $i+1$: If there exists a $B \in K_i^{-1}$ such that $B \not\subseteq B_j$ $(\forall j : 1 \leq j \leq k)$, then by Algorithm 2.8 we compute an A_{i+1}, where $A_{i+1} \in H$, $A_{i+1} \subseteq B$. Set $K(i+1) = K(i) \cup A_{i+1}$. In the converse case we set $H = K(i)$.

Proposition 2.12. [16] *The time complexity of Algorithm 2.11 is*

$$O\left(n(\sum_{q=1}^{m-1}(kl_q + nt_q u_q) + k^2 + n)\right),$$

where $|R| = n$, $|K| = k$, $|H| = m$, *for meanings of* l_q, t_q, u_q *see Proposition 2.6.*

Clearly, in cases for which $l_q \leq k$ $(\forall q \quad 1 \leq q \leq m-1)$ the time complexity of our algorithm is $O(|R|^2|K|^2|H|)$. It is easy to see that in these cases Algorithm 2.11 finds the set of minimal keys in polynomial time in the size of R, K, H. If $|H|$ is polynomial in $|R|$ and $|K|$, then our algorithm is effective. It can be seen that if the number of elements of H is small then Algorithm 2.11 is very effective.

Algorithm 2.13.

Input: Let r be a BCNF relation over R

Output: a BCNF relation scheme s such that $K_s = K_r$

Step 1: From r compute E_r

Step 2: From E_r compute the maximal equality system T_r

Step 3: By Algorithm 2.11 we construct a set of all minimal keys H of r

Step 4: Denoting elements of H by $A_1, \ldots, A_m$ we construct a relation scheme as follows: $s = \langle R, F \rangle$, where $F = \{A_1 \to R, \ldots, A_m \to R\}$.

Based on Theorem 1.10, Algorithm 2.11 and Definition 2.2 we have $K_s = K_r$. It is clear that the time complexity of this algorithm is the time complexity of Algorithm 2.11. In many cases this algorithm is very effective (see Proposition 2.12).

Theorem 2.14. *[14] Let K a Sperner system over R. Denote* $s(K) = \min\{m : |r| = m, K_r = K\}$. *Then* $(2|K^{-1}|)^{1/2} \leq s(K) \leq |K^{-1}| + 1$.

Remark 2.15. Let us take a partition $R = X_1 \cup, \ldots, \cup X_m \cup W$, where $|R| = n$, $m = [n/3]$, and $|X_i| = 3$ $(1 \leq i \leq m)$. We set

$H = \{B : |B| = 2, B \subseteq X_i \text{ for some } i\}$ if $|W| = 0$,

$H = \{B : |B| = 2, B \subseteq X_i \text{ for some } i : 1 \leq i \leq m-1 \text{ or } B \subseteq X_m \cup W\}$ if $|W| = 1$,

$H = \{B : |B| = 2, B \subseteq X_i \text{ for some } i : 1 \leq i \leq m \text{ or } B = W\}$ if $|W| = 2$.

It is easy to see that

$H^{-1} = \{A : |A \cap X_i| = 1, \forall i\}$ if $|W| = 0$,

$H^{-1} = \{A : |A \cap X_i| = 1, (1 \leq i \leq m-1)$ and $|A \cap (X_m \cup W)| = 1\}$ if $|W| = 1$,

$H^{-1} = \{A : |A \cap X_i| = 1, (1 \leq i \leq m)$ and $|A \cap W| = 1\}$ if $|W| = 2$.

If set $K = H^{-1^{-1}}$, i.e. H^{-1} is a set of minimal keys of K, then we have

$K = \{C : |C| = n-3,\ C \cap X_i = \emptyset$ for some $i\}$ if $|W| = 0$,

$K = \{C : |C| = n-3,\ C \cap X_i = \emptyset$ for some i $(1 \leq i \leq m-1)$ or $|C| = n-4, C \cap (X_m \cup W) = \emptyset\}$ if $|W| = 1$,

$K = \{C : |C| = n-3, C \cap X_i = \emptyset$ for some i $(1 \leq i \leq m)$ of $|C| = n-2, C \cap W = \emptyset\}$ if $|W| = 2$.

It is clear that $n-1 \leq |H| \leq n+2$, $3^{[n/4]} < |H^{-1}|$, $|K| \leq m+1$. Based on this partition, Theorem 2.14 and Algorithms 2.7, 2.13 we obtain the following theorems.

Theorem 2.16. *In the BCNF class of relations and relation schemes the time complexity of finding a relation r from a given relation scheme s such that r represents s is exponential in the size of s.*

Proof. We have to prove that:

(1) There is an algorithm finding a BCNF relation r from a given BCNF relation scheme s such that r represents s and the time complexity of this algorithm is exponential time in the size of s.

(2) There exists a BCNF relation scheme $s = \langle R, F\rangle$ such that the number of rows of any BCNF relation representing s is exponential in the size of s.

For (1): we have Algorithm 2.7.

For (2): According to Theorem 2.14 we have $(2|K^{-1}|)^{1/2} \leq s(K)$. We construct a k-relation scheme $s = \langle R, F\rangle$, where $F = \{B \to R : B \in H\}$. It is obvious that $H^{-1} = K_s^{-1}$. Hence, $(2^{1/2}3^{[n/8]}) \leq s(K_s)$ holds. It can be seen that BCNF relation r that is constructed in Algorithm 2.7 has the number of rows at most $|U||H^{-1}| + 1$. Thus, we always can construct a BCNF relation scheme s such that the number of rows of any BCNF relation representing s is exponential in the size of s. The proof of complete.

Theorem 2.17. *In BCNF class of relations and relation schemes over R, the time complexity of finding a relation scheme s from a given relation r such that $K_r = K_s$ is exponential in the number of attributes.*

Proof. It is clear that the worst-case time complexity of Algorithm 2.13 is exponential in the size of R. In Remark 2.15 we have $|K| \leq m+1$. We set $M = \{C - a : \forall a, C : a \in R,\ C \in K\}$. Denote elements of M by $C_1, \ldots, C_t$. Construct a relation $r = \{h_0, h_1, \ldots, h_t\}$ as follows: for all $a \in R$ $h_0(a) = 0$,

for $i = 1, \ldots, t$ $h_i(a) = 0$ if $a \in C_i$, in the converse case $h_i(a) = i$. Clearly, $|r| \leq (m+1)|R| + 1$ holds. We construct a relation scheme $s = \langle R, F \rangle$ with $F = \{A \to R : A \in H^{-1}\}$. It is obvious that $3^{[n/4]} < |F|$ and $K_r = K_s$. Clearly, a minimum cover of any BCNF relation scheme is a k-relation scheme. Thus, we always can construct a BCNF relation r in which the number of rows of r is at most $(m+1)|R| + 1$, but for any BCNF relation scheme $s = \langle R, F \rangle$ such that $K_r = K_s$, the number of elements of F is exponential in the number of attributes. Our proof is complete.

It is known that in the BCNF class a relation r represents a relation scheme s iff r is an Armstrong relation of s. Consequently, our two algorithms and two problems are still true when r is an Armstrong relation of s.

Proposition 2.18. *Let $s = \langle R, F \rangle$ be a relation scheme. Then s is in BCNF iff $\forall B \in K_s^{-1}, a \in B : (B - a)^+ = B - a$.*

Proof. It is easy to see that if s is in BCNF then $(B - a)^+ = B - a$ for $B \in K_s^{-1}$ and $a \in B$. Conversely, assume that s is not in BCNF. Consequently, there is an $A \to \{a\} \in F^+$, where $A^+ \neq R$ and $a \notin A$. By Proposition 2.1 there is a $B \in K_s^{-1}$ such that $A^+ \subseteq B$. Clearly, $a \in B$ and $A \subseteq B - a$. Hence, $(B - a)^+ = B$ holds. The proof is complete.

Proposition 2.18 was independently discovered in [17].

Definition 2.19. Let K be a Sperner system over R. We say that K is unique if K uniquely determines a relation scheme $s = \langle R, F \rangle$, i.e. for every relation scheme $s' = \langle R, F' \rangle$ such that $K_{s'} = K$ we have $F^+ = F'^+$.

It is easy to see that s is a BCNF relation scheme.

It is known [6] that for given a Sperner system K there exists a relation scheme s (a relation r, respectively) such that $K_s = K$ ($K_r = K$, respectively). We say that s (r, respectively) is unique if K_s (K_r, respectively) uniquely determines s (r, respectively), i.e. K_s (K_r, respectively) is unique.

Now we give a necessary and sufficient condition for a given relation scheme to be unique.

Theorem 2.20. *Let $s = \langle R, F \rangle$ be a relation scheme over R. Then s is unique iff for all $a \in A$, $A \in K_s^{-1} : A - a = \cap\{B \in K_s^{-1} : (A - a) \subset B\}$ holds.*

Proof. It is known [5] that a Sperner system K is unique iff for all $B \subseteq A$, $A \in K^{-1}$, B is an intersection of antikeys. Denote $P_s = \{A - a : A \in \in K_s^{-1}, a \in A\}$.

It can be seen that if $s = \langle R, F \rangle$ is unique then $B \in P_s$ implies B is an intersection of antikeys, i.e. $B = \cap\{A \in K_s^{-1} : B \subseteq A\}$. Conversely, assume that for every $B \in P_s$ we have

$$(*) \qquad B = \cap\{A \in K_s^{-1} : B \subseteq A\}.$$

By Proposition 2.18 and according to Proposition 2.1 we have $N_s \subseteq$ $\subseteq (P_s \cup K_s^{-1})$. It can be seen that s is in BCNF. Based on definition of N_s and Proposition 2.1 $K_s^{-1} \subseteq N_s$ holds. According to (*) we obtain $K_s^{-1} = N_s$. Because s is in BCNF we can see that for all $B \subseteq A, A \in K_s^{-1} : B^+ = B$ holds. Thus, B is an intersection of antikeys of s. The proof is complete.

By a polynomial time algorithm finding a set of all antikeys of a given relation (see [14]) and according to Theorem 2.20 we obtain the following proposition.

Proposition 2.21. *There exists an algorithm deciding whether a given relation r is unique. The time complexity of this algorithm is polynomial in the sizes of R and r.*

Theorem 2.20 immediately implies

Corollary 2.22. *Let K be a Sperner system over R. Then there exists a polynomial time algorithm deciding whether a Sperner system H is unique, where $H^{-1} = K$.*

The size of minimal Armstrong relations was investigated in some papers (see [2,5,6,12,14]). Now we present some new bounds for the size of minimal Armstrong relations.

Definition 2.23. (minimal Arsmtrong relation) Let F be an f-family over R. Let $M(F) = \min\{m : |r| = m, F_r = F\}$.

Denote $H_F(A) = \{a \in R : (A, \{a\}) \in F\}$ and $Z(F) = \{A : H_F(A) = A\}$.

Proposition 2.24. [6] *Let F be an f-family over R. Then $(2|N(F)|)^{1/2} <$ $< M(F) \leq |N(F)| + 1$, where $N(F)$ is the minimal generator of $Z(F)$.*

According to the definition of unique Sperner system and based on Theorem 2.20 we have the following

Proposition 2.25. *Let $s = \langle R, F\rangle$ be a relation scheme. Then if K_s is unique then*

$$(2|K_s^{-1}|)^{1/2} < M(F^+) \leq |K_s^{-1}| + 1.$$

Theorem 2.26. *Let K be a Sperner system, $s = \langle R, F\rangle$ a BCNF relation scheme over R. Let*

$M(F^+) = \min\{m : |r| = m, F_r = F^+, s = \langle R, F\rangle$ a relation scheme$\}$;

$K^* = K^{-1} \cup \{B - a : B \in K^{-1}, a \in B, B - a \neq \cap\{C : C \in K^{-1}, B - a \subset$ $\subset C\}\}$.

Then if $K_s = K$ then

$$(*) \qquad (2|K^*|)^{1/2} < M(F^+) \leq |K^*| + 1.$$

Proof. For an arbitrary Sperner system K we can construct a BCNF relation scheme $s = \langle R, F = \{A \to R : A \in K\}\rangle$ such that $K_s = K$. According to the definition of BCNF relation schemes we can see that for all BCNF relation schemes $s' = \langle R, F'\rangle$ such that $K_{s'} = K$ $F^+ = F'^+$ holds. Denoting the elements of K^* by $A_1, \ldots, A_t$ we construct a relation $r = \{h_0, h_1, \ldots, h_t\}$ as follows:

for all $a \in R$, $h_0(a) = 0$, $\forall i = 1, \ldots, t$

$$h_i(a) = \begin{cases} 0 & \text{if } a \in A_i, \\ i & \text{otherwise.} \end{cases}$$

By Proposition 2.18 we obtain $K_r = K$ and r is in BCNF. According to the definition of BCNF we have $F_r = F^+$. Hence, $N(F_r) = N(F^+)$ holds. According to Theorem 1.12 we have $N(F_r) \subseteq E_r$. Based on Proposition 2.18 and the construction of K^*, and $N(F_r) = \{B \in E_r : B \neq \cap\{B' \in E_r : B \subset B'\}\}$, we obtain $N(F_r) = K^*$. According to Proposition 2.24 we have $(2|K^*|)^{1/2} < M(F^+) \leq |K^*| + 1$. The proof is complete.

Let K be a Sperner system over R. We say that an attribute a is prime of K if it belongs to an element of K and nonprime otherwise.

Based on Proposition 2.25 and Theorem 2.26 we obtain the following

Proposition 2.27. *Let K be a Sperner system over $R = \{a_1, \ldots, a_n\}$. Let $K^{-1} = \{A_1, \ldots, A_k\}$ and $K_n = \{a_{i1}, \ldots, a_{is}\}$ be a set of nonprime attributes of K. Then for every BCNF relation scheme $s = \langle R, F\rangle$ such that $K_s = K$*

$$(**) \qquad k(s+1) \leq |N(F^+)| \leq kn.$$

Proof. Clearly, if $B \in K^{-1}$ then $|B| \leq n - 1$. Hence, in Theorem 2.26 $|K^*| \leq kn$. By the proof of Theorem 2.26 $K^* = N(F^+)$ holds. Thus, $N(F^+) \leq kn$ holds. It is known [7] that K_n is the intersection of elements of K^{-1}. According to definition of minimal generator we can see that if $B \in K^{-1}$ then for every $a \in K_n : B - a \in N(F^+)$. Clearly, if $B \in K^{-1}$ then $B \in N(F^+)$. Consequently, we have $k(s+1) \leq |N(F^+)| \leq kn$. The proof is complete.

According to Theorem 2.26 and Proposition 2.27 we can see that for all BCNF relation schemes $s = \langle R, F\rangle$ such that $K_s = K$, $|K^{-1}|$ is polynomial in the number of attributes if and only if the size of minimal Armstrong relation and the number of elements of the minimal generator $N(F^+)$ (sometimes it is denoted by $GEN(s)$) of s also polynomial in $|R|$.

It can be seen that the bounds (*) and (**) are especially interesting when the number of antikeys of K is polynomial in the number of attributes.

Acknowledgments. The authors are grateful to Dr.Uhrin Béla for useful comments to the first version of the manuscript.

References

[1] **Armstrong W.W.,** Dependency structures of database relationships, *Information Processing,* **74** (1974), 580-583.

[2] **Beeri C., Dowd M., Fagin R. and Staman R.,** On the structure of Armstrong relations for functional dependencies, *J. ACM,* **31** (1) (1984), 30-46.

[3] **Beeri C. and Bernstein P.A.,** Computational problems related to the design of normal form relational schemas, *ACM Trans. on Database Syst.,* **4** (1) (1979), 30-59.

[4] **Bernstein P.A.,** Synthesizing third normal form relations from functional dependencies, *ACM Trans. on Database Syst.,* **1** (1976), 227-298.

[5] **Burosch G., Demetrovics J. and Katona G.O.H.,** The poset of closures as a model of changing databases, *Order,* **4** (1987), 127-142.

[6] **Demetrovics J.,** Logical and structural investigation of relation datamodel, *MTA SZTAKI Tanulmányok,* **114** (1980), 1-97. (in Hungarian)

[7] **Demetrovics J. and Thi V.D.,** Some results about functional dependencies, *Acta Cybernetica,* **8** (3) (1988), 273-278.

[8] **Demetrovics J. and Thi V.D.,** Relations and minimal keys, *Acta Cybernetica,* **8** (3) (1988), 279-285.

[9] **Demetrovics J. and Thi V.D.,** On keys in the relational datamodel, *EIK* **24** (10) (1988), 515-519.

[10] **Gottlob G. and Libkin L.,** Investigations on Armstrong relations, dependency inference and excluded functional dependencies, *Acta Cybernetica,* **9** (4) (1990), 385-402.

[11] **Lucchesi C.L. and Osborn S.L.,** Candidate keys for relations, *J.Comput. Syst. Sci.,* **17** (2) (1978), 270-279.

[12] **Mannila H. and Raiha K.J.,** Design by example: An application of Armstrong relations, *J.Comput. Syst. Sci.,* **33** (1986), 126-141.

[13] **Mannila H. and Raiha K.J.,** Algorithms for inferring functional dependencies, *Proc. of the 13th Int. Conf. on Very Large Data Bases, Brighton, September 1987.*

[14] **Thi V.D.,** *Investigation on combinatorial characterizations related to functional dependency in the relational datamodel,* Ph.D. dissertation, *MTA SZTAKI Tanulmányok,* **191** (1986), 1-157. (in Hungarian)

[15] **Thi V.D.,** Minimal keys and antikeys, *Acta Cybernetica,* **7** (4) (1986), 361-671.

[16] **Thi V.D.,** On antikeys in the relational datamodel, *Alk. Mat. Lapok*, **12** (1986), 111-124. (in Hungarian)

[17] **Demetrovics J., Hencsey G., Libkin L.O. and Muchnik I.B.,** Normal form relation schemes: a new characterization, *Acta Cybernetica,* **10** (3) (1992), 141-153.

(Received September 22, 1994)

J. Demetrovics and Vu Duc Thi
Computer and Automation Institute
Hungarian Academy of Sciences
P.O.B. 63
H-1502 Budapest, Hungary