CS 348: Assignment 4(3)

Question 1.
Each of the following parts are questions on normal forms and their computation.

1. Exhibit a sequence of binary lossless join decompositions of relation schema $R/(A, B, C, D, E)$ that obtains a decomposition for which each relation is in BCNF, assuming the FDs $\{A \to BC, CD \to E, B \to D, E \to A\}$ hold on R. Prove that each relation in your decomposition is indeed in BCNF.

R is not in BCNF because of FD $B \to D$, which is in F (and therefore in F^+), is non-trivial, and for which B is not a superkey of R.

A single iteration of the while loop of $ComputeBCNF(R, F)$ therefore obtains a binary decomposition of R into relations

$$R_1/(B, D) \quad \text{and} \quad R_2/(A, B, C, E).$$

To check that each R_i is in BCNF, one needs to ensure no FD $X \to Y$ exists over R_i that is non-trivial, is in F^+, and for which $X \to R_i$ is not in F^+, that is, has a left hand side X that is not a superkey of R_i. In the case of R_2, there are $2^{|R_2|} - 1$ possibilities for X and likewise for Y, which means checking 225 possible FDs! This can be pruned in two ways.

1. By the union and decomposition rules, $X \to Y$ is a non-trivial FD in F^+ where $X \to R_2$ is not in F^+ if and only if this also holds for one of the FDs obtained by substituting Y with one of its members. Thus, one only needs to consider FDs where Y is a single attribute.

2. To satisfy the conditions that $X \to Y$ over R_2, where Y is a single attribute, is non-trivial and that $X \to R_2$ is not in F^+, one only needs to consider where $|X| \leq |R_2| - 2$ since any non-trivial FD with a larger X cannot disqualify R_2 from satisfying BCNF.

Thus, there are only 24 non-trivial FDs over R_2 with one or two attributes on the left hand side and one attribute on the right hand side that need to be checked. This can be done exhaustively, or perhaps more quickly by first computing the candidate keys of R_2:

$$\{A, BC, E\},$$

and then checking that there are no non-trivial FDs over R_2 in F^+ with either just B or just C on the left hand side (again, by using *Compute*X^+).

By this reasoning, note that any relation schema consisting of just two attributes must be in BCNF; thus R_1 is also in BCNF.

---

2. Exhibit a lossless join and dependency preserving decomposition of relation schema $R/(A, B, C, D, E, F)$ by algorithm *Compute*3NF, assuming the FDs $\{A \to BCD, BC \to DE, B \to D, D \to A\}$ hold on R. Note that this requires showing (a) a minimal cover, and (b) a proof that one of the relations in the decomposition contains a candidate key.

The first part of *Compute*3NF requires computing either of the following minimal covers for the FDs (depending on whether or not one applies "Step 4"):

$$\{\ B \to D,\ D \to A,\ A \to B,\ A \to C,\ B \to E\ \} \quad \text{or} \quad \{\ B \to DE,\ D \to A,\ A \to BC\ \}$$

This leads to the following initial (part of a) decomposition of R:

$$
\begin{array}{lll}
\{\ BD, & & \\
\quad DA, & & \{\ BDE, \\
\quad AB, & \text{or} & \quad DA, \\
\quad AC, & & \quad ABC \quad \} \\
\quad BE \quad \} & &
\end{array}
$$

However, none of the relation schemata deriving from an FD in any minimal cover contains a candidate key. If one did, say R_i, then R_i would qualify as a superkey and $R_i \to R$ would be a logical consequence of either of the minimal covers, which none are. Thus, the second part of *Compute*3NF requires computing a candidate key and adding this to complete the 3NF decomposition. The set of attributes AF is a candidate key since $AF \to R$ in a logical consequence of either of the minimal covers, but neither $A \to R$ nor $F \to R$ is, all of which follows by using $ComputeX^+$.

Thus, lossless join and dependency preserving decompositions are given by either of the following:

$$
\begin{array}{lll}
\{\ BD, & & \\
\quad DA, & & \{\ BDE, \\
\quad AB, & \text{or} & \quad DA, \\
\quad AC, & & \quad ABC, \\
\quad BE & & \quad AF \quad \}. \\
\quad AF \quad \} & &
\end{array}
$$