APPENDIX

A. NP-completeness of Lossless Extraction Decision Problem

We define the decision version L of lossless extraction problem as follows.

Input: $(\mathcal{F}, V)$ and k, where $\mathcal{F}$ is the set of all the frequent patterns such that every $f \in \mathcal{F}$ has a valid branch combination b_f with storage cost c_f, V is the universal set of all the branch ids of the frequent patterns, and $k \in \mathbb{N}$.

Question: Does there exist a $E \subseteq \mathcal{F}$ for lossless extraction with $\sum\limits_{e \in E} c_e \leq k$?

Claim 1: L is in **NP**.

Proof: The following verifier for L runs in time polynomial in the length of the inputs:

Verifier $\mathbb{V}$ $(< \mathcal{F}, V, k >, < E >)$:

1. Construct $U = \{u | u \in X \wedge X \in \mathcal{F}\}$
2. Construct $T = \{e | e \in X \wedge X \in E\}$
3. $\forall e \in U$, Construct $V_e = \{i | e \in X \wedge X \in \mathcal{F} \wedge i \in b_X\}$
4. $\forall e \in T$, Construct $W_e = \{i | e \in X \wedge X \in E \wedge i \in b_X\}$
5. If the following are all true then accept else reject:

   i $U = T$ (all frequent items included)

   ii $\forall e \in U, V_e = W_e$ (all <branch id, item> included)

   iii $\sum\limits_{e \in E} c_e \leq k$ (total cost $\leq k$) ■

Claim 2: *Weighted Set Cover* $\leq_p L$

Proof: The weighted set cover problem takes as input $< U, \mathcal{A}, k >$, where $\mathcal{A}$ is a set whose member $i \subseteq U$ with cost c_i and $k \in \mathbb{N}$, to answer the question if U has an $\mathcal{A}$-cover of cost k.

Let us define a function f that takes as input $< U, \mathcal{A}, k >$ and output $< \mathcal{F}, V, k >$. We perform this by computing mutually exclusive sets D_i from $\mathcal{A}$ in polynomial time, such that $D_i \cap D_j = \emptyset$, $\forall i \neq j$ and $\bigcup_i D_i = U$, where $i \in \mathbb{N}$.

Now, $\mathcal{F}$ can be computed as follows.
$\forall w \in \mathcal{A} \implies w \in \mathcal{F}$ with cost c_w and valid branch combination $\{i\}$ where $w \subseteq D_i$.

Suppose U has a weighted $\mathcal{A}$-cover $\mathcal{B}$ of cost $\leq k$. Then it can be trivially shown, $\mathcal{B}$ can be used for lossless extraction of $\mathcal{F}$, since all frequent items and their corresponding valid branches are covered.

Conversely, if $\mathcal{B}$ can be used for lossless extraction of $\mathcal{F}$, then $\bigcup_{w \in \mathcal{B}} w = U$ and total cost of $\mathcal{B} \leq k$. Thus $\mathcal{B}$ is also a weighted $\mathcal{A}$-cover of cost $\leq k$. ■