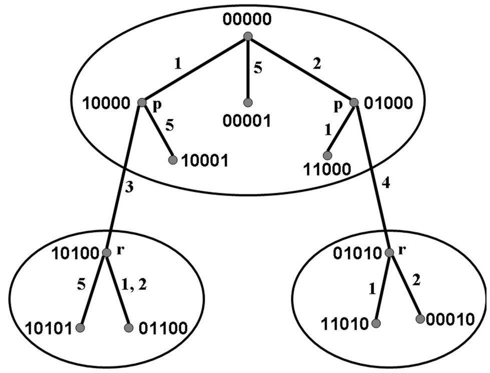

Fig. 5. Example illustrating the reconstruction. Underlying phylogeny is T_I^*; taxa r and p (both could be Steiner) are guessed to create $E = \{(10000, 10100), (01000, 01010)\}$; E induces three components in T_I^*. When all taxa in T_I^* are considered, character 3 conflicts with $1, 2$ and 5 and character 4 conflicts with 1 and 2; two components are perfect (penalty 0) and one has penalty 2; $\texttt{penalty}(I) =_{def} \texttt{penalty}(T_I^*) = 7$.

B. Correctness

We will now prove the correctness of the pseudo-code under the assumption that all the guesses performed by our algorithm are correct. Specifically, we will show that if $\texttt{penalty}(I) \leq q$ then function `buildNPP` returns an optimum phylogeny. The following lemma proves the correctness of our algorithm.

Lemma 4.1: At any point in execution of the algorithm, an optimum phylogeny for I can be constructed as $E \cup (\cup_i T_i)$, where T_i is *any* optimum phylogeny for $M_i \in L$.

Proof: We prove the lemma using induction. The lemma is clearly true at the beginning of the routine when $L = \{I\}, E = \emptyset$. As inductive hypothesis, assume that the above property is true right before an execution of Step 2e. Consider any optimum phylogeny $T_{M_j}^*$ where $c(v)$ mutates exactly once and on the edge (r, p). Phylogeny $T_{M_j}^*$ can be decomposed into $T_{M_j}^*(c(v), 0) \cup T_{M_j}^*(c(v), 1) \cup (r, p)$ with length $l = \texttt{length}(T_{M_j}^*(c(v), 0)) + \texttt{length}(T_{M_j}^*(c(v), 1)) + d(r, p)$. Again, since $c(v)$ mutates exactly once in $T_{M_j}^*$, all the taxa in $M0$ and $M1$ are also in $T_{M_j}^*(c(v), 0)$ and $T_{M_j}^*(c(v), 1)$ respectively. Let T', T'' be *arbitrary* optimum phylogenies for $M0$ and $M1$ respectively. Since $p \in M0$ and $r \in M1$ we know that $T' \cup T'' \cup (r, p)$ is a phylogeny for M_j with cost $\texttt{length}(T') + \texttt{length}(T'') + d(r, p) \leq l$. By the inductive hypothesis, we know that an optimum phylogeny for I can be constructed using any optimum phylogeny for M_j. We have now shown that using any optimum phylogeny for $M0$ and $M1$ and adding edge (r, p) we can