constraint. We note that the Algorithm II can build a set using an optional greedy stage. This set is further refined in the enhanced local search stage comprising of addition, deletion and swap operations. At the termination of the LS stage we obtain the primary choice $\breve{\underline{\mathcal{G}}}$. Then, both stages are repeated over the complement set, $\underline{\Omega} \setminus \breve{\underline{\mathcal{G}}}$ to generate an alternate choice, $\tilde{\underline{\mathcal{G}}}$. Finally, the choice yielding the larger weighted sum rate utility among the primary and alternate choices is chosen. Regarding the slight variation alluded to above, we note that the direct adaptation would have initialized the Enhanced LS stage with the empty set or the singleton set yielding the highest weighted rate, since the LS stage also includes adding (or insertion) of elements. However, in our numerical simulations we saw that initializing the LS stage using the output of the greedy stage helps in reducing the run-time without performance degradation. We proceed to derive performance guarantee for Algorithm II. Towards that end, we introduce an assumption pertaining to the feasibility of the minimum rates. We emphasize that this assumption is only for deriving a performance guarantee and is not needed for implementing the algorithm.

• *Admission control assumption: Each macro TP $m \in \mathcal{M}$ can itself simultaneously meet twice the minimum rates of all users u that are present in at-least one tuple $(u, b) \in \underline{\Omega}$ for any $b \in \mathcal{B}_m$.*

We now offer the following result which holds even when the greedy stage is skipped and which assumes that Algorithm II is initialized with $\mathrm{MaxIter} = \infty$ and $\Delta = \frac{\epsilon}{C}$, where $\epsilon > 0$ and C is a large enough constant that depends (polynomially) on the size of $\underline{\Omega}$.

Theorem 3. *Algorithm II yields a constant factor ($\frac{1}{4+\epsilon}$) approximation to (1) over all input instances for which the admission control assumption holds.*

Proof. Let $\tilde{\underline{\mathcal{I}}}$ denote the family of sets obtained by taking the pairwise union of members of $\underline{\mathcal{I}}$. We define an extended set function as $\tilde{f}^{\mathrm{wsr}}(\underline{\mathcal{G}}) = \sum_{m \in \mathcal{M}} \tilde{f}^{\mathrm{wsr}}_m(\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}),\ \forall\ \underline{\mathcal{G}} \in \tilde{\underline{\mathcal{I}}}$. Here, for any $\underline{\mathcal{G}} \in \tilde{\underline{\mathcal{I}}}$ we define $\tilde{f}^{\mathrm{wsr}}_m(\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}) = \hat{O}(1, \mathbf{1})$, where $\hat{O}(1, \mathbf{1})$ is computed as described in Section III-A for the macro TP m and its set of pico TPs $\mathcal{B}_m$ using unit budgets and the given association in $\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$, with the following caveat. In particular, now in obtaining the user sets $\{\mathcal{U}^{(b)}\}$ we treat the user in each tuple $(u, b) \in \underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$ as a distinct virtual user. Hence, if (u, b_1) and (u, b_2) belong to $\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$, we suppose that two distinct virtual users with their own separate peak rates and associated maximum and minimum rate limits are specified. These peak rates and limits are