Subsequence automata with default transitions

Let $S$ be a string of length $n$ with characters from an alphabet of size $\sigma$. The subsequence automaton of $S$ (often called the directed acyclic subsequence graph) is the minimal deterministic finite automaton accepting all subsequences of $S$. A straightforward construction shows that the size (number of states and transitions) of the subsequence automaton is $O(n\sigma)$ and that this bound is asymptotically optimal. In this paper, we consider subsequence automata with default transitions, that is, special transitions to be taken only if none of the regular transitions match the current character, and which do not consume the current character. We show that with default transitions, much smaller subsequence automata are possible, and provide a full trade-off between the size of the automaton and the delay, i.e., the maximum number of consecutive default transitions followed before consuming a character. Specifically, given any integer parameter $k$, $1 < k \leq \sigma$, we present a subsequence automaton with default transitions of size $O(nk\log \sigma)$ and delay $O(\log \sigma)$. Hence, with $k = 2$ we obtain an automaton of size $O(n \log \sigma)$ and delay $O(\log \sigma)$. At the other extreme, with $k = \sigma$, we obtain an automaton of size $O(n \sigma)$ and delay $O(1)$, thus matching the bound for the standard subsequence automaton construction. Finally, we generalize the result to multiple strings. The key component of our result is a novel hierarchical automata construction of independent interest.