CS 374 Spring 2018
Homework 2

Nathaniel Murphy (njmurph3@illinois.edu)
Tanvi Modi (tmodi3@illinois.edu)
Marianne Huang (mhuang46@illinois.edu)

Problem 2

Let us first define a few helper functions.

- Define $b_k : 2^k \to \mathcal{P}(\Sigma_k)$ such that for a given binary string s of length k, $b_k(s) = \{a \in \Sigma_k \mid s_a = 1\}$. Notice that $b_k(s) \subseteq \Sigma_k$.
- Define $h_k : \{0, 1\} \to \Sigma_k$, where

$$h_k(i) = \begin{cases} \epsilon & \text{if } i = 0 \\ k & \text{if } i = 1 \end{cases}$$

- Define $c_b : 2^k \to \Sigma_k^*$ inductively.

$$c_k(w) = \begin{cases} \epsilon & \text{if } w = \epsilon \\ h_{|k|}(a) \cdot c_{|k-1|}(u) & \text{if } w = au \end{cases}$$

The language requires that we remember if a symbol is encountered before we read the $\natural$ symbol so that we can process all symbols correctly after the $\natural$ symbol. Let us prove this statement.

Let us create a fooling set

$$F = \{c_k(s) \mid s \in 2^k\}$$

Choosing arbitrary $i, j \in F \setminus \{0^k\}, i \neq j$. Without loss of generality, let us assume that $|i| \geq |j|$ (swap them if necessary). We see that the set $\big(b_k(i) \setminus b_k(j)\big)$ must be nonempty because no symbols in i or j repeat within each string and $|i| \geq |j|$. Let $w = \natural \cdot a,\ a \in \big(b_k(i) \setminus b_k(j)\big)$. We see that $iw = i\natural a \in T_k$, but $jw = j\natural a \notin T_k$ because $a \in b_k(i)$, but $a \notin b_k(j)$.

$|F| = 2^k$, so we see that M, the DFA representing T_k must have at least 2^k elements.

Let us now prove a stronger statement. Denote:

$$L_b = \{c_k(s) \mid s \in 2^k\} \text{ and } L_{b\natural} = \{c_k(s) \cdot \natural \mid s \in 2^k\}$$

Let $F' = L_b \cup L_{b\natural}$. We have 4 cases for any i, j chosen from F':

Case $i, j \in L_b,\ i \neq j$: Without loss of generality, let $|i| \geq |j|$ (swap them if necessary). Above, we have chosen a $w = \natural \cdot a \in \big(b_k(i) \setminus b_k(j)\big)$ such that $iw \in T_k$ and $jw \notin T_k$.

Case $i \in L_b, j \in L_{b\natural}$: Let $j = j'\natural$. Let $w = \natural i$. It is clear that $iw = i\natural i \in T_k$, but $jw = j'\natural\natural i \notin T_k$.

Case $i \in L_{b\natural}, j \in L_b$: Same case as above, but reverse i nad j.

Case $i, j \in L_{b\natural}$: Notice that i, j can be written in the form $i'\natural, j'\natural$. Without loss of generality, let $|i'| \geq |j'|$ (swap them if necessary). We see that the set $\big(b_k(i') \setminus b_k(j')\big)$ must be nonempty because no symbols in i or j repeat within each string and $|i'| \geq |j'|$. Choose $w = a \in \big(b_k(i') \setminus b_k(j')\big)$. It follows that $iw = i'\natural a \in T_k$ because $a \in b_k(i')$, while $jw = j'\natural a \notin T_k$ because $a \notin b_k(j')$.

We have been able to produce a $w \in \Sigma_k^* \cup \{\natural\}$ such that $\forall\, i, j \in \Sigma_k^* \cup \{\natural\}$, either $iw \in T_k \wedge jw \notin T_k$ or $iw \notin T_k \wedge jw \in T_k$.

Notice that $F' = L_b \cup L_{b\natural}$. and $L_b \cap L_{b\natural} = \emptyset$ trivially, $|F'| = |L_b| + |L_{b\natural}| = 2^k + 2^k = 2 \cdot 2^k = 2^{k+1}$, so we now see that M, the DFA representing the language T_k must have at least 2^{k+1} states.
$\square$