Name: Rollno:

CSE340: Theory of Computation (Homework Assignment 2)

Due Date: 12th September, 2017 (in class)

Total Number of Pages: 4 Total Points 40

Question 1. (5 points) Give a regular expression for the following language.

$$B = \{x \in \{0,1\}^* \mid x \text{ does not contain the substring } 101\}$$

Solution:

$$0^*(1^*000^*)^*1^*0^*$$

The expression in bracket encodes the fact that 1 is either followed by a 1 or a 00. The initial 0^* and the last 1^*0^* cover the boundary cases.

Question 2. (5 points) Prove that $\{0^i1^j \mid gcd(i,j) = 1\}$ is not regular?

Solution: We prove this using the pumping lemma for regular languages. Let p_i be the ith prime number.

Given k, choose $w = 0^{p_{k+1}}1^{p_1p_2\ldots p_k}$. Note that $gcd(p_{k+1}, p_1p_2\ldots p_k) = 1$ for all $k \geq 1$. Now given partition, $w = xyz$, such that $|xy| \leq k$ and $|y| \geq 1$, observe that x and y consists of 0's only. Let $|y| = l$. Now by choosing $i = 0$ we have $xy^0z = 0^{p_{k+1}-l}1^{p_1p_2\ldots p_k}$.

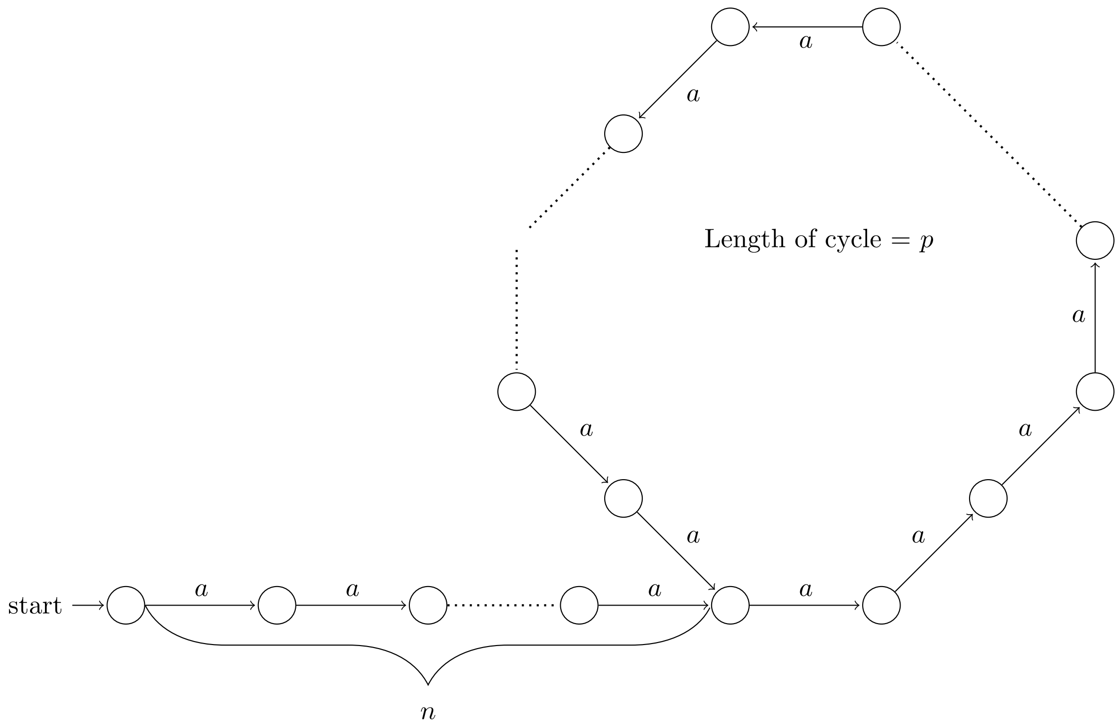
Observe that any positive integer strictly smaller than a prime number (in this case p_{k+1}) must have some prime factor p_i such that $p_i \leq p_k$. Therefore $gcd(p_{k+1} - l, p_1p_2\ldots p_k) > 1$ and thus xy^0z is not in the given language.

Question 3. (6 points) Let $A \subseteq \mathbb{N}$ be a subset of natural numbers. A is said to be *ultimately periodic* if there exists numbers $p > 0$ and $n \geq 0$, such that for all $m \geq n$, $m \in A$ if and only if $m + p \in A$. In other words, after a certain point (the number n) the numbers in the set A occur in a fixed regular interval of length p.

Consider $L \subseteq \{a\}^*$. Prove that L is regular if and only if the set $\{m \mid a^m \in L\}$ is ultimately periodic.

(Hint: Think how will the DFA of a unary regular language look like.)

Solution: Suppose L is regular and let M be a DFA for L. Since every state in M has exactly one outgoing transition, M will have the following structure.



Now set n to be the length of the initial "non-cyclic" portion and p be the length of the cycle in M. Now if a string a^m such that $m \geq n$, terminates at a state q then a^{m+p} will also terminate at the state q. Therefore $a^m \in L$ then a^{m+p} is also in L.

For the other direction, suppose the set $\{m \mid a^m \in L\}$ is ultimately periodic. Let $n \geq 0$ and $p > 0$ be the constants as per the definition of ultimate periodicity. Construct a DFA $M_{n,p}$ which has an initial linear portion of length n, followed by a cycle of length p as shown in the above figure. For all strings $w' \in L$ of length at most $n + p$, make the state which is at a distance $|w'|$ from the start state, an accept state. Clearly this DFA accepts all strings in L.

Question 4. (12 points) Give CFGs for the following languages

(a) $L_1 = \{a^i b^j c^k d^l \mid i, j, k, l \geq 1,\ i = l,\ j = k\}$

Solution:

$$
\begin{aligned}
S &\longrightarrow aSd \mid aTd \\
T &\longrightarrow bTc \mid bc
\end{aligned}
$$

Important point to be noted here is that a string contains at least one a, b, c, d.

(b) $L_2 = \{a^n b^m \mid n, m \geq 0,\ n \neq m\}$

Solution: We divide the language into two parts: strings where $n > m$ and strings where

$n < m$ and design a grammar accordingly.

$$\begin{aligned} S &\longrightarrow AT \mid TB \\ T &\longrightarrow aTb \mid \epsilon \\ A &\longrightarrow aA \mid a \\ B &\longrightarrow bB \mid b \end{aligned}$$

(c) $L_3 = \{a^i b^j c^k \mid i, j, k \geq 0,\ i > j \text{ or } j > k\}$

Solution: Once again we consider the two cases differently: $i > j$ and $j > k$.

$$\begin{aligned} S &\longrightarrow AaT_1C \mid ABbT_2 \\ T_1 &\longrightarrow aT_1b \mid \epsilon \\ T_2 &\longrightarrow bT_2c \mid \epsilon \\ A &\longrightarrow aA \mid \epsilon \\ B &\longrightarrow bB \mid \epsilon \\ C &\longrightarrow cC \mid \epsilon \end{aligned}$$

Question 5. Consider the following CFG G over the set of terminals $T = \{+, *, 0, 1, (,)\}$

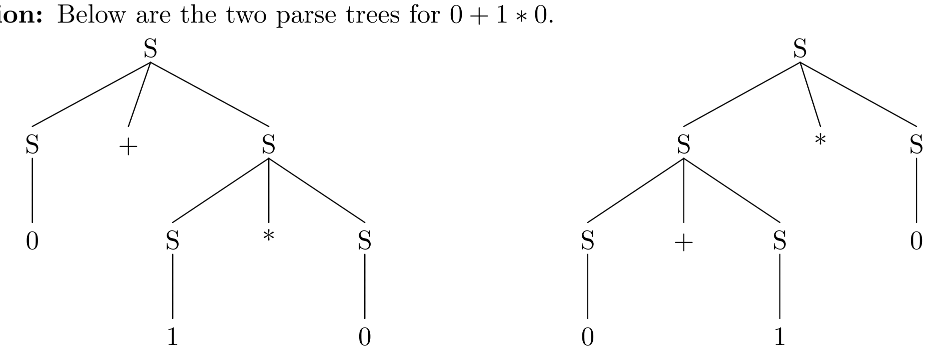
$$S \longrightarrow S + S \mid S * S \mid (S) \mid 0 \mid 1$$

(a) (2 points) Give a string of length 5 that is ambiguous with respect to G.

Solution: $0 + 1 * 0$

(b) (4 points) Give two parse trees for the string in part (a) with respect to G.

Solution: Below are the two parse trees for $0 + 1 * 0$.



(c) (6 points) Give an unambiguous CFG for the language generated by the above grammar that gives proper precedence to the operators (i.e. highest precedence to brackets followed by the $*$ operator and then the $+$ operator).

Solution:

$$
\begin{array}{rcl}
S & \longrightarrow & S + A \mid A \\
A & \longrightarrow & A * B \mid B \\
B & \longrightarrow & (S) \mid 0 \mid 1
\end{array}
$$