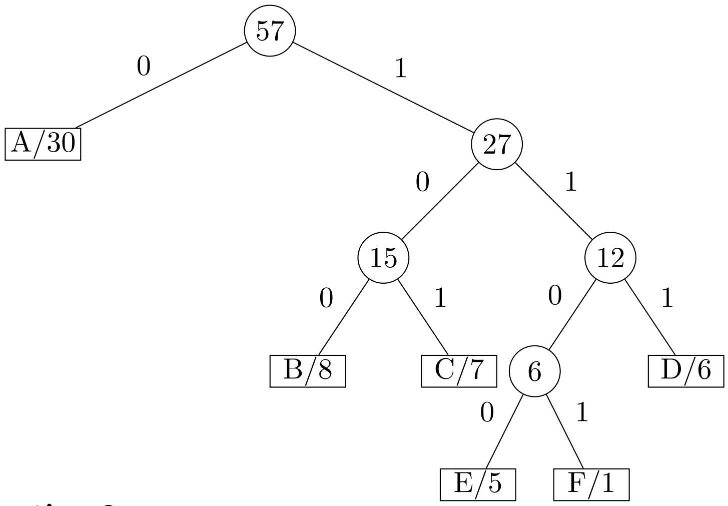
COMP S265F Design and Analysis of Algorithms
Lab 4: Huffman Codes – Suggested Solution

Question 1. Steps to construct the Huffman code tree:

1. Merge E & F to (E,F).
2. Merge D & (E,F) to (D,E,F).
3. Merge B & C to (B,C).
4. Merge (B,C) & (D,E,F) to (B,C,D,E,F).
5. Merge A & (B,C,D,E,F) to (A,B,C,D,E,F).



Char	Huffman code
A	0
B	100
C	101
D	111
E	1100
F	1101

Question 2.

(a) We can find the minimum number of gas stops by the greedy strategy that Keith only stops at a gas station s_i if he does not have enough gas to go to the next station s_{i+1}.

$S \leftarrow \emptyset$ ▷ S is the set of gas stations Keith should stop
$d \leftarrow d_0$ ▷ d is the distance traveled from the last gas stop
for $i \leftarrow 1$ to k **do**
  $d \leftarrow d + d_i$
  if $d > n$ **then** ▷ if Keith doesn't stop at s_i, he does not have enough gas to s_{i+1}
    $S \leftarrow S \cup \{s_i\}$ ▷ Thus, Keith must stop at s_i
    $d \leftarrow d_i$
  end if
end for
return S

(b) We can show that our solution is optimal, i.e., S contains the minimum number of gas stops, using proof by contradiction.

Let $t = |S|$, i.e., the number of gas stops in our solution. Suppose, for the sake of contradiction, that the optimal solution has less than t stops. Let $S = \{s_{g_1}, s_{g_2}, \ldots, s_{g_t}\}$. Then, we can divide all the k gas stations in the highway into $t+1$ sets:

- $\{1, 2, \ldots, g_1\}$
- $\{g_1 + 1, g_1 + 2, \ldots, g_2\}$
- $\{g_2 + 1, g_2 + 2, \ldots, g_3\}$
- $\ldots$
- $\{g_{t-1} + 1, g_{t-1} + 2, \ldots, g_t\}$

- $\{g_t + 1, g_t + 2, \ldots, k\}$

Consider the first t sets. Since the optimal solution has less than t stops, there exists i such that the optimal solution does not stop at any gas station in the i-th set. For convenience, if $i = 1$, let $g_0 = 0$. The i-th set contains the gas stations $g_{i-1} + 1, g_2 + 2, \ldots, g_i$.

Recall that in our solution, Keith stops at the gas stations g_{i-1} and g_i. According to our greedy algorithm, we choose g_i to stop because $d_{g_{i-1}} + d_{g_{i-1}+1} + d_{g_{i-1}+2} + \cdots + d_{g_i} > n$. Thus, the distance between the gas stations g_{i-1} and $g_i + 1$ is larger than n.

In the optimal solution, Keith's bike will not have enough fuel to travel from gas station g_{i-1} to gas station $g_i + 1$, which contradicts that it is a valid solution. Therefore, there does not exist any solution with less stops than our solution.