1. (20 points) Prove, calculate or rank the asymptotic bounds for the following functions.

   (a) (10 points) Given the n-th harmonic number $H(n)$:

   $$H(n) = \sum_{k=1}^{n} 1/k \quad (n > 0),$$

   prove $H(n) = O(\log n)$ and $H(n) = \Omega(\log n)$.

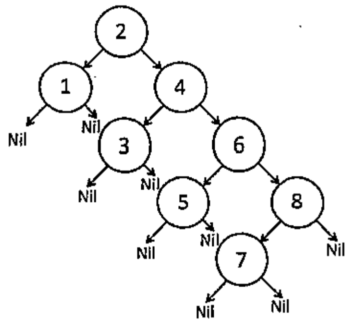
   (b) (5 points) Give the asymptotically tight bound (θ−notation) for the recurrence $T(n)$:

   $$T(n) = 2T(n/2) + n/\log n \quad (assume\ T(1) = 1).$$

   Show the derivation.

   (c) (5 points) Rank the following functions $f_1 \dots f_5$ by order of growth.
   If $f_i = O(f_j)$ $(i \neq j)$, then $f_i \prec f_j$.
   $f_1 = \log^2 n, \quad f_2 = n \log n, \quad f_3 = \sqrt{n},$
   $f_4 = \log(n!), \quad f_5 = (3/2)^n$
   Please complete the form: ___ $\prec$ ___ $\prec$ ___ $\prec$ ___ $\prec$ ___.

2. (10 points) A red-black tree is a special type of binary search trees. Given the following binary search tree on 8 keys (i.e., integers 1 … 8), please answer the questions.



   (a) (5 points) The tree cannot be simply transformed to a red-black tree by coloring its internal nodes. Explain what property of the red-black tree cannot be satisfied.

   (b) (5 points) Explain how to transform this tree to a red-black tree through a single rotation. Draw the constructed red-black tree and label its nodes with either B (black) or R (red) color.

3. (17 points) Huffman coding is a type of coding schemes used for lossless data compression. In the coding scheme, each character in a file is represented by a unique string, which is called a codeword. The length of a codeword that represents a frequently appeared character in the input file is usually shorter than other codewords as well as the original character. Thus, when the entire file is encoded by converting all characters in the file to codewords which represent them, the total length can usually be significantly reduced. In a binary Huffman code, the number of symbols that can be used to form a codeword is 2. We can also use a n-ary Huffman code, in which the number of symbols that can be used to form a codeword is n.

   (a) (8 points) Please prove that the decoding tree which represents an optimal code for a file (using which the encoded file has a minimum length) must be a full n-ary tree, i.e., all non-leaf nodes in the tree have n children.

   (b) (4 points) You are given the numbers of appearances of all characters in a file as follows: $\{'a' : 700,\ 'b' : 400,\ 'c' : 200,\ 'd' : 100,\ 'e' : 1300,\ 'f' : 2400,\ 'g' : 100\}$. What are the lengths of the codewords for the characters 'a' and 'c', respectively, in a 3-ary Huffman code derived for this file?

(c) (5 points) Which of the following statements are correct? Please mark all correct statements.

i. A Huffman code is a fixed-length code.

ii. A Huffman code is a prefix code.

iii. If a file contains a sequence of 8-bit characters such that the maximum character frequency is less than **three times** of the minimum character frequency, then binary Huffman codings always cannot do better than an ordinary 8-bit fixed-length code.

iv. If a file contains a sequence of 8-bit characters such that the maximum character frequency is less than **twice** of the minimum character frequency, then binary Huffman codings always cannot do better than an ordinary 8-bit fixed-length code.

v. If a file contains a sequence of 8-bit characters such that the maximum character frequency is less than **1.5 times** of the minimum character frequency, then binary Huffman codings always cannot do better than an ordinary 8-bit fixed-length code.

4. (10 points) You are given a **convex polygon** $P = \{v_1, v_2, \ldots, v_n\}$, where $v_i, 1 \leq i \leq n$, are the vertices that form the polygon. In a convex polygon, each internal angle is less than or equal to 180 degrees and each line segment between two vertices remains inside or on the boundary of the polygon.

A triangulation of a convex polygon is a set of non-intersecting diagonals that partitions the polygon into triangles. We define that the weight of a triangulation is the sum of the lengths of the perimeters of all component triangles. The minimum triangulation problem is to find the triangulation of the given polygon that has the minimum weight.

Denote the cost of a minimum triangulation of the polygon formed by $\{v_i, v_{i+1}, \ldots, v_j\}$ as $c(i, j)$. Then the minimum triangulation has a weight of $c(1, n)$.

(a) (6 points) Fill the blanks below to complete the recurrences which can be used for a dynamic programming algorithm to solve this problem.

$$c(i,j) = \begin{cases} 0 & ,\underline{\quad\text{(a)}\quad} \\ \min_{i<k<j}\{\underline{\quad\text{(b)}\quad}\} & ,\text{otherwise} \end{cases}$$

(b) (4 points) Using the recurrences shown in the previous subproblem to develop a dynamic programming algorithm to solve the problem as efficient as possible, what would be its time complexity in terms of n, the number of vertices in the given convex polygon P?

5. (13 points) Suppose that we use disjoint-set forests with union by rank and path compression to represent disjoint sets.

(a) (6 points) Fill the blanks in the following programs that perform the three operations for disjoint sets.

```
Make-Set(x)
   x.parent=x
   x.rank=0

Union(x,y)
   Link(Find-Set(x),Find-Set(y))

Link(x,y)
   if ___(a)___
      y.parent=x
   else
      x.parent=y
      if ___(b)___
         y.rank=y.rank+1
```

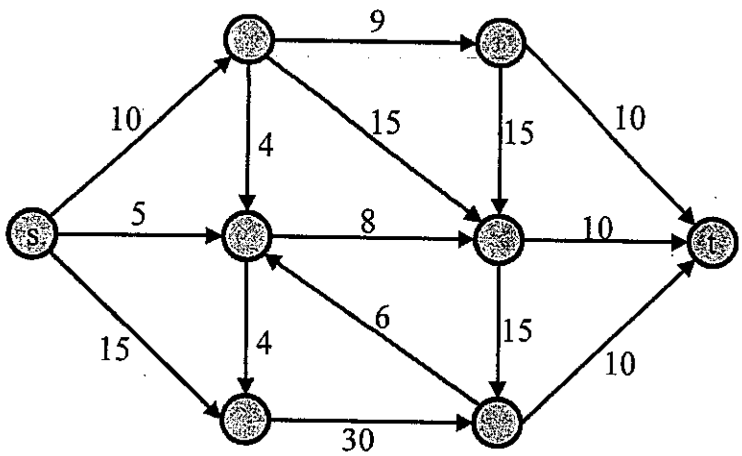
接次頁

```
Find-Set(x)
   if x≠x.parent
        (c)
   return x.parent
```

(b) (7 points) Draw the data structure that results after running the following program. For each tree node, please mark its parent and labels its identification number (in the form of x_i, $1 \le i \le 16$)

```
for i=1 to 16
   Make-Set(x_i)
for i=1 to 15 by 2
   Union(x_i, x_{i+1})
for i=1 to 13 by 4
   Union(x_i, x_{i+2})
Union(x_1, x_15)
Union(x_11, x_13)
Union(x_1, x_10)
Find-Set(x_2)
Find-Set(x_9)
```

6. (15 points) What is the value of minimum cut of the following flow network? What is the residual network corresponding to your solution?



7. (15 points) 3-SAT and SUBSET-SUM are two NP-complete problems. In 3-SAT, we are asked whether a given boolean formula ϕ is satisfiable, where ϕ is in 3-conjunctive normal form (3-CNF). In the SUBSET-SUM problem, we are given a finite set of integers, S, and a target integer t. We ask whether there is a subset $S' \subseteq S$ whose elements sum to t. It has been shown that 3-SAT is polynomial time reducible to SUBSET-SUM. Given the following boolean formula in 3-CNF,

$$\phi = (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (\overline{x_1} \vee \overline{x_2} \vee \overline{x_3})$$

what is its corresponding SUBSET-SUM problem? (You need to show your construction process.)

試題隨卷繳回