Heap

The following heap property is satisfied:

- for any two nodes $i$, $j$ such that $i$ is the parent of $j$, we have $\text{key}[A[i]] \leq \text{key}[A[j]]$.

A heap. Numbers in the circles denote key values of elements.
insert($v, key\_value$)
insert\((v, key\_value)\)
insert\((v, key\_value)\)
insert\((v, key\_value)\)
$\text{insert}(v, \text{key_value})$
**insert**\((v, key\_value)\)

1: \( s \leftarrow s + 1 \)
2: \( A[s] \leftarrow v \)
3: \( p[v] \leftarrow s \)
4: \( key[v] \leftarrow key\_value \)
5: \( \text{heapify\_up}\(s)\)

**heapify-up\(i)\**

1: \( \text{while } i > 1 \text{ do} \)
2: \( j \leftarrow \lfloor i/2 \rfloor \)
3: \( \text{if } key[A[i]] < key[A[j]] \text{ then} \)
4: \( \text{swap } A[i] \text{ and } A[j] \)
5: \( p[A[i]] \leftarrow i, p[A[j]] \leftarrow j \)
6: \( i \leftarrow j \)
7: \( \text{else break} \)
extract_min()
extract_min()
extract_min()
extract_min()
extract_min()
extract_min()
extract_min()
1: ret ← A[1]
3: p[A[1]] ← 1
4: s ← s − 1
5: if s ≥ 1 then
6:    heapify_down(1)
7: return ret

decrease_key(v, key_value)
1: key[v] ← key_value
2: heapify-up(p[v])

heapify-down(i)
1: while 2i ≤ s do
2:    if 2i = s or
3:       key[A[2i]] ≤ key[A[2i + 1]] then
4:       j ← 2i
5:    else
6:        j ← 2i + 1
7: if key[A[j]] < key[A[i]] then
8:    swap A[i] and A[j]
10:   else break
Running time of heapify\textsubscript{up} and heapify\textsubscript{down}: $O(\lg n)$
- Running time of heapify\_up and heapify\_down: $O(lg\ n)$
- Running time of insert, exact\_min and decrease\_key: $O(lg\ n)$
- Running time of heapify\_up and heapify\_down: $O(lg\ n)$
- Running time of insert, exact\_min and decrease\_key: $O(lg\ n)$

<table>
<thead>
<tr>
<th>data structures</th>
<th>insert</th>
<th>extract_min</th>
<th>decrease_key</th>
</tr>
</thead>
<tbody>
<tr>
<td>array</td>
<td>$O(1)$</td>
<td>$O(n)$</td>
<td>$O(1)$</td>
</tr>
<tr>
<td>sorted array</td>
<td>$O(n)$</td>
<td>$O(1)$</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>heap</td>
<td>$O(lg\ n)$</td>
<td>$O(lg\ n)$</td>
<td>$O(lg\ n)$</td>
</tr>
</tbody>
</table>
Two Definitions Needed to Prove that the Procedures Maintain **Heap Property**

**Def.** We say that $H$ is almost a heap except that $key[A[i]]$ is too small if we can increase $key[A[i]]$ to make $H$ a heap.

**Def.** We say that $H$ is almost a heap except that $key[A[i]]$ is too big if we can decrease $key[A[i]]$ to make $H$ a heap.
Outline

1. Toy Example: Box Packing
2. Interval Scheduling
   • Interval Partitioning
3. Offline Caching
   • Heap: Concrete Data Structure for Priority Queue
4. Data Compression and Huffman Code
5. Summary
6. Exercise Problems
Encoding Letters Using Bits

- 8 letters $a, b, c, d, e, f, g, h$ in a language
- need to encode a message using bits
- idea: use 3 bits per letter

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>000</td>
<td>001</td>
<td>010</td>
<td>011</td>
<td>100</td>
<td>101</td>
<td>110</td>
<td>111</td>
</tr>
</tbody>
</table>

\[ deacfg \rightarrow 011100000010101110 \]

Q: Can we have a better encoding scheme?

- Seems unlikely: must use 3 bits per letter

Q: What if some letters appear more frequently than the others?
Q: If some letters appear more frequently than the others, can we have a better encoding scheme?

A: Using variable-length encoding scheme might be more efficient.

Idea

- using fewer bits for letters that are more frequently used, and more bits for letters that are less frequently used.
Q: What is the issue with the following encoding scheme?

- $a: 0$
- $b: 1$
- $c: 00$

A: Can not guarantee a unique decoding. For example, 00 can be decoded to $aa$ or $c$.

Solution: Use prefix codes to guarantee a unique decoding.
Q: What is the issue with the following encoding scheme?

- $a$: 0
- $b$: 1
- $c$: 00

A: Can not guarantee a unique decoding. For example, 00 can be decoded to $aa$ or $c$. Use prefix codes to guarantee a unique decoding.
Q: What is the issue with the following encoding scheme?
- $a$: 0
- $b$: 1
- $c$: 00

A: Can not guarantee a unique decoding. For example, 00 can be decoded to $aa$ or $c$.

Solution
Use prefix codes to guarantee a unique decoding.
Prefix Codes

**Def.** A prefix code for a set $S$ of letters is a function $\gamma : S \rightarrow \{0, 1\}^*$ such that for two distinct $x, y \in S$, $\gamma(x)$ is not a prefix of $\gamma(y)$. 
**Def.** A prefix code for a set $S$ of letters is a function $\gamma : S \rightarrow \{0, 1\}^*$ such that for two distinct $x, y \in S$, $\gamma(x)$ is not a prefix of $\gamma(y)$.

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
<td></td>
</tr>
</tbody>
</table>
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 00010011000000001011110100001001
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001100000001011110100001001

- c
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

0001/001/100000001011110100001001

ca
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/000001011110100001001
- cad
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
<td></td>
</tr>
</tbody>
</table>

- 001/001/100/0000/0101110100001001
- cadb
Prefix Codes Guarantee Unique Decoding

Reason: there is only one way to cut the first code.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

0001/001/100/0000/01/011110100001001

cadbh
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/0000/01/01/1110100001001
- cadbh
Prefix Codes Guarantee Unique Decoding

- **Reason:** there is only one way to cut the first code.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
<td>g</td>
<td>h</td>
</tr>
<tr>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/0000/01/01/11/10100001001
- cadbhhe
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>001</th>
<th>0000</th>
<th>0001</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/0000/01/01/11/1010/0001001
- cadbhhef
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/0000/01/01/11/1010/0001/001
- cadbhhef\text{c}
Prefix Codes Guarantee Unique Decoding

- Reason: there is only one way to cut the first code.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>001</td>
<td>0000</td>
<td>0001</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td>11</td>
<td>1010</td>
<td>1011</td>
<td>01</td>
</tr>
</tbody>
</table>

- 0001/001/100/0000/01/01/11/1010/0001/001/
- cadbhhefca
Properties of Encoding Tree

A rooted binary tree with left edges labelled 0 and right edges labelled 1. Nodes that correspond to codes for some letter. A non-leaf has exactly two children if the coding scheme is not wasteful.

Best Prefix Codes

Input: frequencies of letters in a message
Output: prefix coding scheme with the shortest encoding for the message
Properties of Encoding Tree

- Rooted binary tree

The diagram shows a rooted binary tree with labels on the edges indicating 0 and 1. The nodes correspond to letters, and the path from the root to each node represents a binary code for that letter.
Properties of Encoding Tree

- Rooted binary tree
- Left edges labelled 0 and right edges labelled 1
Properties of Encoding Tree

- Rooted binary tree
- Left edges labelled 0 and right edges labelled 1
- A leaf corresponds to a code for some letter

Input:
- frequencies of letters in a message

Output:
- prefix coding scheme with the shortest encoding for the message
**Properties of Encoding Tree**

- Rooted binary tree
- Left edges labelled 0 and right edges labelled 1
- A leaf corresponds to a code for some letter
- If coding scheme is not wasteful: a non-leaf has exactly two children
Properties of Encoding Tree
- Rooted binary tree
- Left edges labelled 0 and right edges labelled 1
- A leaf corresponds to a code for some letter
- If coding scheme is not wasteful: a non-leaf has exactly two children

Best Prefix Codes
- **Input:** frequencies of letters in a message
- **Output:** prefix coding scheme with the shortest encoding for the message
### Example

<table>
<thead>
<tr>
<th>letters</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequencies</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

| scheme 1 length | 2 | 3 | 3 | 2 | 2 |
| total = 89 |

| scheme 2 length | 1 | 3 | 3 | 3 | 3 |
| total = 87 |

| scheme 3 length | 1 | 4 | 4 | 3 | 2 |
| total = 84 |

- **Scheme 1**
  - a → d → e
  - b → c

- **Scheme 2**
  - a
  - b → c
  - d → e

- **Scheme 3**
  - a
  - e
  - d
  - b → c
### Example

<table>
<thead>
<tr>
<th>letters</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequencies</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>scheme 1 length</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>total</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scheme 2 length</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>total</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scheme 3 length</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>total</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Scheme 1**
```
  a
 / 
b   d
 / 
|   e
b   c
```

**Scheme 2**
```
  a
 / 
|   b
 / 
c   d
 / 
|   e
```

**Scheme 3**
```
  a
 / 
|   e
 / 
|   d
 / 
|   b
```

<table>
<thead>
<tr>
<th>Scheme 1</th>
<th>Scheme 2</th>
<th>Scheme 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Scheme 1" /></td>
<td><img src="image2.png" alt="Scheme 2" /></td>
<td><img src="image3.png" alt="Scheme 3" /></td>
</tr>
</tbody>
</table>
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

**Q:** What types of decisions should we make?
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

**Q:** What types of decisions should we make?

- Can we directly give a code for some letter?
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

Q: What types of decisions should we make?

- Can we directly give a code for some letter?
- Hard to design a strategy; residual problem is complicated.
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

Q: What types of decisions should we make?

- Can we directly give a code for some letter?
- Hard to design a strategy; residual problem is complicated.
- Can we partition the letters into left and right sub-trees?
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

**Q:** What types of decisions should we make?

- Can we directly give a code for some letter?
- Hard to design a strategy; residual problem is complicated.
- Can we partition the letters into left and right sub-trees?
- Not clear how to design the greedy algorithm
Example Input: \((a: 18, b: 3, c: 4, d: 6, e: 10)\)

**Q:** What types of decisions should we make?

- Can we directly give a code for some letter?
- Hard to design a strategy; residual problem is complicated.
- Can we partition the letters into left and right sub-trees?
- Not clear how to design the greedy algorithm

**A:** We can choose two letters and make them brothers in the tree.
Which Two Letters Can Be Safely Put Together As Brothers?

- Focus on the “structure” of the optimum encoding tree