



1.4.2 Hash Tables

Q1

Andrew is writing code to implement an allow list for a firewall to keep a track of websites that users on a network are permitted to visit. He has decided to store the address of each website, along with any extra details, in a hash table.

The hash table's hash function is carried out on the website's address which acts as the key.

The hash function works in the following way:

1. Discard the character up to and including the first dot.
2. Discard the characters including and to the right of the remaining leftmost dot.
3. Convert the characters to uppercase.
4. Add the ASCII values of the characters together and MOD by 100.

A	B	C	D	E	F	G	H	I	J	K	L	M
65	66	67	68	69	70	71	72	73	74	75	76	77
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
78	79	80	81	82	83	84	85	86	87	88	89	90

For example, `www.python.org.uk` gets hashed in the following manner:

Step 1:

`python.org.uk`

Step 2:

`python`

Step 3:

`PYTHON`

Step 4:

$(80 + 89 + 84 + 72 + 79 + 78) \text{ MOD } 100 = 82$

a) State what hashed value would be given by the website `www.pugs.co.uk`

[1 mark]





b) Write a function hash which takes a string and returns the hashed value.

You can assume you have access to the following three functions:

- `asc()` – this takes a character as a parameter and returns its ASCII value.
- `locate()` – this takes a string and a character as parameters and returns the index of the first instance of the character in the string.
- `upper()` – this takes a string as a parameter and returns the strings with all characters in uppercase.

[5 marks]

c) The hash function that Andrew is using is producing many instances where two different keys generate the same hash value after the hash function is applied to them. This causes collisions.

Explain how collision handling methods can be applied to handle collisions.

[4 marks]

The hash function is changed so there are no longer high numbers of collisions.

During busy periods the firewall is expected to check several addresses a second. It is anticipated that roughly 10 new addresses will be added to the allows list each day.

Andrew is considering whether a hash table (with the new hash function) is the best approach, or if the whitelist would be better stored in a linked list.

d) Discuss whether a hash table or linked list is better to store acceptable websites. You should compare how each structure can be searched and has data added and come to a recommendation as to which is better for the allow list.

[12 marks]

