Edwards, Ryan
399067099

Assignment 4

CS 320
September 28, 2019

Theoretical questions

1. Illustrate that if a new email from the address $a \in A$, it always gets through (1 pts).

   Proof. Consider the email address: redw764@aucklanduni.ac.nz

   In binary, this is equivalent to:

   01110010 01100101 01100100 01110111 00110111 00110110 00110100 01000000 01100001 01110101 01100011 01101011 01101100 01100001 01101110 01100100 01110101 01101110 01101001 00101110 01100001 01100011 00101110 01101110 01111010

   Let this be represented as an integer x such that $x \in A$

   When the hash table is constructed, the function will set $B[h(x)] = 1$

   When an email is received, the email filter will apply the hash function $h(x)$

   By definition:

   - If $B[h(x)] = 1$ the email will go through
   - If $B[h(x)] = 0$ the email is considered spam

   Because $B[h(x)] = 1$ was set when the hash table was constructed, and because a hash function will always return the same output for $h(x)$, the lookup function will return 1 and the email will go through.

   □

2. Given any position $0 \leq i < n$, what is the probability that $B[i] = 1$ (2 pts).

   As there is a universal hashing function for integers

   $$h_{ab}(x) = ((ax + b) \mod p) \mod n$$

   As $n = 8,000,000,000$ the probability of a collision is $Pr_h[h(x) = h(y)] \leq \frac{1}{n} \leq \frac{1}{8\text{B}}$

   Therefore, the probability of $B[i] = 0$ after 1 insert is $\geq 1 - \frac{1}{8\text{B}}$

   The probability of $B[i] = 0$ after 1B inserts is $\geq (1 - \frac{1}{8\text{B}})^{1\text{B}}$

   Thus, the the probability that $B[i] = 1$ is simply

   $$\leq 1 - (1 - \frac{1}{8\text{B}})^{1\text{B}} \lessapprox 0.11750$$

3. Given a spam email from the address $a' \notin A$, what is the probability that it gets through (2 pts)

   As, by definition, a universal hashing function will uniformly distribute a new hashed value, the probability of collision is the probability that $B[i] = 1$.

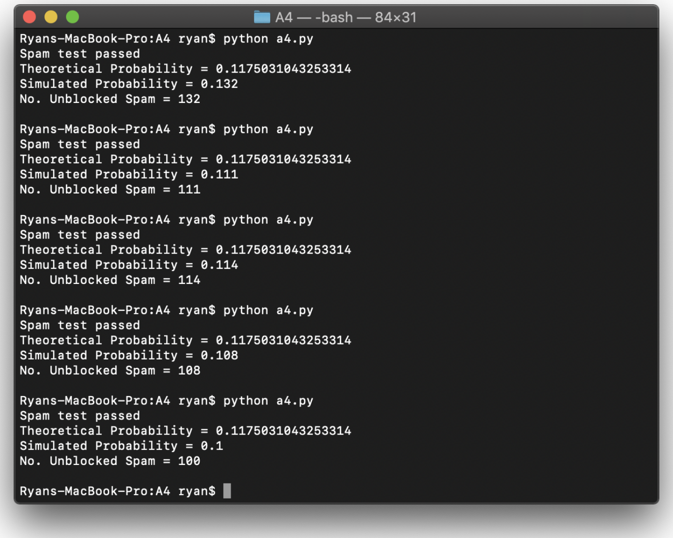
   Thus, the probability of $B[h(a')] = 1$ is $\lessapprox 0.11750$

Practical implementation

```
import random

def main():
    # Initial Values
    n = 8000000
    p = 8024047 # prime number > n

    # Initalise an empty Hash Table
    hash_table = [0] * n

    # 0 <= a, b < p
    a = random.randrange(0, p)
    b = random.randrange(0, p)

    # Create the email address list
    total_addresses = 1000000
    email_address_list = [i for i in range(1, total_addresses + 1)]

    for address in email_address_list:
        hash_table[universal_hash(a, b, n, p, address)] = 1


### Question one ###
    try:
        for number in email_address_list:
            hash_value = universal_hash(a , b, n, p, number)

            if hash_table[hash_value] == 0:
                raise SpamDetected

    except SpamDetected:
        print("Spam␣test␣failed")

    else:
        print("Spam␣test␣passed")

### Question two ###
    # Based on formula in theoretical question 2
    theoretical_probability = 1 - (1 - 1/n) ** total_addresses

    print("Theoretical␣Probability␣=", theoretical_probability)
```

```
### Question three ###
    spam_email_count = 0
    spam_email_no = 1000

    for i in range(spam_email_no):
        random_address = random.randrange(total_addresses + 1, 9999999)
        hash_value = universal_hash(a, b, n, p, random_address)
        if hash_table[hash_value] == 1:
            spam_email_count += 1

    print("Simulated Probability =", spam_email_count / spam_email_no)
    print("No. Unblocked Spam =", spam_email_count)
    print()


# Hashing function based on universal hash family for integer
def universal_hash(a, b, n,  p, x):
    return ((a * x + b) % p) % n

class Error(Exception):
   """Base class for other exceptions"""
   pass
class SpamDetected(Error):
   """Spam has been detected"""
   pass

main()
```

Sample Output:



```
Ryans-MacBook-Pro:A4 ryan$ python a4.py
Spam test passed
Theoretical Probability = 0.1175031043253314
Simulated Probability = 0.132
No. Unblocked Spam = 132

Ryans-MacBook-Pro:A4 ryan$ python a4.py
Spam test passed
Theoretical Probability = 0.1175031043253314
Simulated Probability = 0.111
No. Unblocked Spam = 111

Ryans-MacBook-Pro:A4 ryan$ python a4.py
Spam test passed
Theoretical Probability = 0.1175031043253314
Simulated Probability = 0.114
No. Unblocked Spam = 114

Ryans-MacBook-Pro:A4 ryan$ python a4.py
Spam test passed
Theoretical Probability = 0.1175031043253314
Simulated Probability = 0.108
No. Unblocked Spam = 108

Ryans-MacBook-Pro:A4 ryan$ python a4.py
Spam test passed
Theoretical Probability = 0.1175031043253314
Simulated Probability = 0.1
No. Unblocked Spam = 100

Ryans-MacBook-Pro:A4 ryan$
```