# Assignment 2

8116775 - Michael McMillan

## Question 1

a. The distribution of the cases requiring each number of hashes.

- **Puzzle A Distribution:**

It shows a uniform distribution where each number of hashes from 1 to 32 appears exactly once. This suggests that for Puzzle A, with 1 sub-puzzle and k = 5, each possible hash value (from the set of 32 hashes) is used exactly once, so the number of cases requiring each number of hashes is always 1.

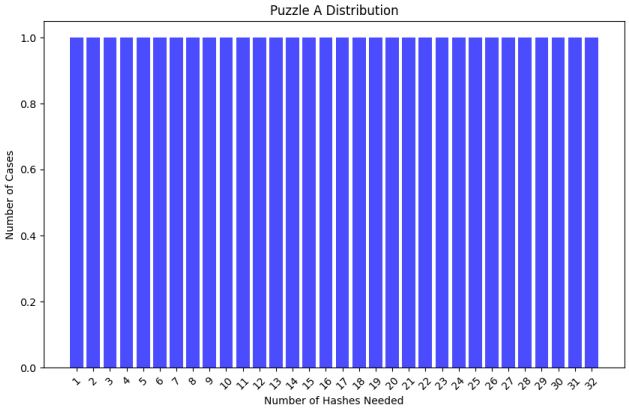
- **Puzzle B Distribution:**

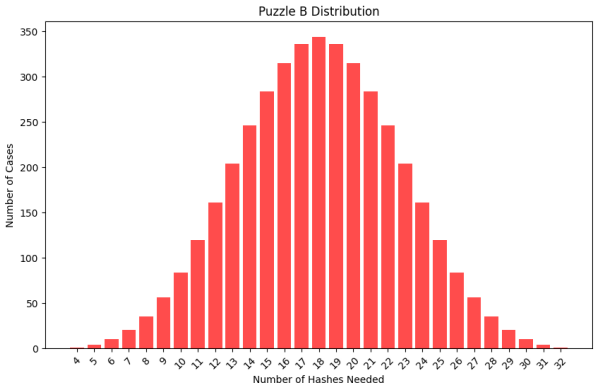
  It resembles a normal distribution, where the number of cases increases to a peak around the centre and then decreases symmetrically. This is expected for Puzzle B with 4 sub-puzzles and k=3. The sum of hash values from combinations of 4 sub-puzzles produces a more diverse range of sums, with most combinations leading to sums around the middle of the possible range.

b. Explain the method you used to obtain your distributions.

I implemented a Python program to simulate the puzzle scenarios for Puzzle A (k = 5) and Puzzle B (k = 3). The program randomly selects a solution within the possible solution space and then iterates through all possible solutions, counting the number of hash checks until the correct solution is found. This process was repeated for 10,000 trials for each puzzle, and the distribution of hash counts was recorded.

c. A graph of the distribution of the data above.




d. The average number of hashes needed.

- Puzzle A: Average Hashes = 16.5
- Puzzle B: Average Hashes = 18

e. The population standard deviation for the distribution of the number of hashes needed.

- **Puzzle A:**
  - Standard Deviation = 9.233

**Puzzle B:**

  - Standard Deviation = 4.583

Student#: 8116775

Student Name: Michael McMillan

# Question 2

Assuming that the TCP SYN packet is 32 bytes in
size. How much bandwidth does the attacker consume to continue this attack?

**2. Sustaining the Attack:**

- The attacker needs to ensure that, after 150 seconds, each slot in the table that gets purged is immediately filled again with a new TCP SYN request.
- Since each connection request stays for 2.5 minutes (150 seconds), the attacker must replenish each of the 512 slots once every 2.5 minutes.

The rate of requests per minute:

$$Rate = \frac{512\ connection\ requests}{2.5\ minutes} = \frac{512}{2.5} = 204.8\ requests\ per\ minutes$$

To keep the table full, the attacker must send 204.8 TCP SYN packets per minute.

**3. Bandwidth Consumption:**

- Each TCP SYN packet is 32 bytes.
- To calculate the bandwidth consumed by the attacker, we need to multiply the number of packets per second by the size of each packet.

Since I have calculated the attacker sends 204.8 rpm, we can convert this to packets per second:

$$Packets\ per\ second = \frac{204.8}{60} \approx 3.413\ bytes\ per\ second$$

The size of each packet is 32 bytes:

$$Bandwidth\ per\ second = \frac{204.8}{60} \approx 109.22\ bytes\ per\ second$$

To convert this to a more standard bandwidth unit (bits per second):

$$Bandwidth\ in\ bps = 109.22 \times 8 = 873.76\ bps\ (bits\ per\ second)$$

Student#: 8116775

Student Name: Michael McMillan

# Question 3

What is the probability that the message is okay (no malware)? Justify your answer using Bayes theorem.

1. **Define Events:**
   - $V$: The message is viral.
   - $\sim V$: The message is not viral.
   - $T$: The malware checker flags the message as viral.
2. **Given Data:**
   - Incidence of viral attachments: $P(V) = \frac{1}{250} = 0.004$
   - Probability of correctly identifying a viral message: $P(T|V) = 0.95$
   - Probability of correctly identifying a non-viral message: $P(\sim T| \sim V) = 0.95$
   - False positive rate $P(\sim T| \sim V) = 1 - P(\sim T| \sim V) = 0.05$
3. **Bayes Theorem:**
   - To find the probability that the message is not viral, given that it has been flagged as viral, compute $P(\sim V|T)$

$$P(\neg V \mid T) = \frac{P(T| \sim V) \times P(\sim V)}{P(T)}$$

4. **Calculate the Total Probability of Flagging as Viral:**

$$P(T) = P(T|V) \times P(V) + P(T| \sim V) \times P(\sim V)$$

$$P(T) = (0.95 \times 0.004) + (0.05 \times 0.996) = 0.0038 + 0.0498 = 0.0536$$

5. **Substitute into Bayes' Theorem:**

$$P(\sim V|T) = \frac{0.05 \times 0.996}{0.0536} \approx 0.929$$

6. **Conclusion:** sss probability that a flagged message is okay (i.e., not malware) is 92.9%. This high probability reflects the low incidence of malware and the malware checker's false positive rate.

# Question 4

What is the main difficulty with using CAPTCHAs?

1. **User Frustration:** CAPTCHAs can be difficult for users to solve, leading to frustration and potentially discouraging legitimate users from completing tasks.
2. **Accessibility Issues:** CAPTCHAs can be problematic for users with disabilities, such as visually impaired users who may have difficulty solving visual CAPTCHAs.
3. **Evolving Bot Technologies:** AI and machine learning advancements make bots increasingly capable of solving CAPTCHAs, reducing their effectiveness over time.
4. **Impact on User Experience:** CAPTCHAs add an extra step to the login or registration process, potentially impacting user experience and conversion rates.
5. **False Positives:** CAPTCHAs may incorrectly identify legitimate users as bots, preventing them from completing tasks.

Student#: 8116775

Student Name: Michael McMillan

# Question 5

What are honeypots? How are they better at resisting spam bots than CAPTCHAs?
A honeypot is a cybersecurity mechanism designed to lure and trap malicious actors, such as attackers or bots, by mimicking legitimate systems or resources. They are typically used to detect, analyse, and counteract cyberattacks by attracting attackers to a decoy system or resource that appears genuine.
Compared to CAPTCHAs, which block bots by asking them to prove they're human, honeypots are better because they don't bother real users and can quietly collect more detailed information on how bots behave. This makes them more effective in resisting spam bots

# Question 6

Briefly describe, in your own words, each of the following. Be sure to specify the domain and nature of each.

a. WannaCry

WannaCry is a ransomware attack that targeted computers running Microsoft Windows in May 2017. It encrypted users' files and demanded a Bitcoin ransom payment to unlock them. The attack spread rapidly through a vulnerability in Windows, affecting thousands of organisations worldwide, including hospitals and businesses.

b. XML Bomb.

An XML bomb is a Denial-of-Service attack that exploits how XML data is processed. It involves creating a maliciously large or complex XML document that, when parsed by an XML processor, consumes excessive system resources, causing the system to crash or become unresponsive.

# Question 7

a. Assume there is no limit to the query size. Give a sequence of two queries that will identify Fran's salary.

- **Query 1: Query for All Female Physics Lecturers**

```
SELECT SUM(Salary)
FROM Employees
WHERE Gender='Female' AND School = 'Physics' AND Position = 'Lecturer';
```

This will give you the salaries of all female physics lecturers, including Fran.

- **Query 2: Query for all Physics Teachers**

```
SELECT SUM(Salary)
FROM Employees
WHERE School = 'Physics' AND Position = 'Lecturer';
```

This will give us the physics lecturers' salaries (including Fran and Ewen's).

- **Query 3: Calculate Fran's Salary**
  - Fran's Salary = (Result of Query 1) – (Result of Query 2 excluding Fran's Salary)
  - Given that Query 1 isolates Fran directly, the result of Query 1 is Fran's salary.

b. Suppose there is a lower and upper query size limit that satisfies $k \leq |X(C)| \leq N - k$ with k = 2. Show a sequence of queries that could be used to determine Diana's salary.

**Step 1: Query to Isolate two specific female lecturers' salaries**

- Query 1: SELECT SUM(Salary) FROM Employees WHERE Gender = 'Female' AND School = 'Computing'

Student#: 8116775

Student Name: Michael McMillan

**Step 2: Query to include all female lecturers**

- Query 2: `SELECT SUM(Salary) FROM Employees WHERE Gender = 'Female';`
  - This will give us the sum of all female employees' salaries, including Diana, Carol, Fran, Kim and Ivana.

**Step 3: Calculate Diana's Salary**

To isolate Diana's salary, subtract Ivana's known salaries (from other queries or logically) from the result of Query 1, then subtract those known salaries from Query 2 and compare these results to find Diana's exact wage.

Student#: 8116775

Student Name: Michael McMillan