School of Computing & Information Technology

CSCI262 System Security
Spring 2024

Assignment 2 (10 marks, worth 10%)

1. You have two puzzles with parameters as follows:

   `Puzzle A:` One sub–puzzle. $k = 5$.

   `Puzzle B:` Four sub-puzzles. $k = 3$.

   You should provide, for both cases other than part (b), the following:

   (a) The distribution of the number of cases that require each number of hashes. **1 Mark**

   (b) Explain the method you used to obtain your distributions. Don't go into too many details or show working, it's more "I wrote a C++ program to ... and then using ... I ...". **0.5 Mark**

   (c) A graph of the distribution of the data above. **0.5 Mark**

   (d) The average number of hashes needed. **0.5 Mark**

   (e) The population standard deviation for the distribution of the number of hashes needed. **0.5 Mark**

   You should assume that if there are N possible solutions you check the N^{th} by hashing even if all others have failed and there has to be a solution.

2. Using a TCP SYN spoofing attack, the attacker aims to flood the table of TCP connection requests on a system so that it is unable to respond to legitimate connection requests. Consider a server system with a table for 512 connection requests. This system will retry sending the SYN-ACK packet five times when it fails to receive an ACK packet in response, at 30 second intervals, before purging the request from its table. Assume that no additional countermeasures are used against this attack and that the attacker has filled this table with an initial flood of connection requests. At what rate (per minute) must the attacker continue to send TCP connection requests to this system in order to ensure that the table remains full? Assuming that the TCP SYN packet is 32 bytes in size. How much bandwidth does the attacker consume to continue this attack? **1 Mark**

3. Consider that the incidence of viral attachments in email messages in 1 in 250. Your malware checker will correctly identify a message as viral 95% of the time. Your malware checker will correctly identify a message as non–viral 95% of the time. Your malware checker has just flagged a message as being malware. What is the probability that the message is actually okay (no malware)? Justify your answer using Bayes theorem. **1 Mark**

4. Many websites use a CAPTCHA image on their login page. A typical application of this is in an HTML form asking for the email ID and the login password of a user. The webpage also shows some numbers and letters, modified in a manner such that it is still easy for a human to recognize these characters. The user is then asked to recognize these characters and is granted login access only when they successfully enter the characters. Explain how using a CAPTCHA can help prevent email spam. What is the main difficulty with using CAPTCHAs?. **1 Marks**

5. What are honeypots? How are they better at resisting spam bots than CAPTCHAs? **1 Mark**

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

   (a) WannaCry. **0.5 Mark**

   (b) XML Bomb. **0.5 Mark**

7. Consider the database below and answer the questions based on it.

Name	Gender	School	Position	Salary
Alex	Male	Computing	Lecturer	\$80,000
Bob	Male	Mathematics	Lecturer	\$60,000
Carol	Female	Mathematics	Lecturer	\$100,000
Diana	Female	Computing	Lecturer	\$65,000
Ewen	Male	Physics	Lecturer	\$72,000
Fran	Female	Physics	Lecturer	\$98,000
Gary	Male	Computing	Administrator	\$40,000
Humphry	Male	Mathematics	Lecturer	\$72,000
Ivana	Female	Computing	Tutor	\$12,000
Jeff	Male	Physics	Administrator	\$80,000
Kim	Female	Mathematics	Lecturer	\$100,000
Lex	Male	Computing	Tutor	\$12,000
Morris	Male	Engineering	Tutor	\$15,000

Assume you only have a statistical interface, so only aggregate queries will be successful. You know Fran is a female Physics Lecturer and Dania is a female Computing lecturer. The questions below explore how we might determine her salary using inference, in the presence of various query size restrictions.

   (a) Assume there is no limit on the query size. Give a sequence of two queries that will identify the salary of Fran. **1 Mark**

   (b) Suppose that 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. **1 Mark**

Notes on submission

1. Submission is via Moodle. Your submission file should be a PDF file.
2. Late submissions will be marked with a 25% deduction for each day, including days over the weekend.
3. Submissions more than three days late will not be marked, unless an extension has been granted.
4. If you need an extension apply through SOLS, if possible **before** the assignment deadline.
5. Plagiarism is treated seriously. Students involved will likely receive zero.