Assignment 1

Instructions: read all the questions carefully and explain your answers with examples and figures where it is required. If you want to reuse a diagram from slides, it is always recommended to redraw and give reference. **This assessment has three parts and total of 100 points, you need to attempt all the questions, please give your opinion, rather than using someone else. The deadline for the assignment is 31st of March 23:59 pm.**

Part 1 (25 Points)

Q1. Confidentially, Integrity and availability are considered as three main components of the information security. Classify each of the element in following two scenarios and explain how they are implemented: (5 points)

1. Web server handling online sales for the computer hardware parts.
2. ATM machines

Q2. Technically, one time pad encryption looks perfect and it's hard to break it; however, one time pad has weak points and its hardly in the practice. Please state the reasons with an example that explain why it's not in practice. What are the streamciphers, why we need pseudorandom generators (PRG) with stream ciphers? Use diagrams where appropriate. (5 points)

Q3. What is the difference between hash functions and message authentication code (MAC)? Why HMAC looks more secure? Justify with an example. Highlight the shortcomings of both mechanisms. Please use figures where it is appropriate. (5 points)

Q4. What is 256 in SHA-256? Why SHA256 is better than the previous versions of SHA like SHA-1? Compare the performance of SHA256 with MD5. Please justify your answer with examples. (5 points)

Question 5: In a university scenario, the administration wants to grant students access to the Canvas learning management system using secret keys. However, the university also wants to maintain control over the creation of these secret keys by keeping one master key. How can the university implement a system to generate secret keys for all students while ensuring the security of the master key? (Hint: each student has a unique and public sid) (5 points)

Part 2 (50 Points)

Q1: Password based authentication mechanism has been involved for several generations. Following questions demonstrate how to measure security for each generation and the weaknesses. Consider a user Alice and an attacker A.

A. Alice chooses a random four-digit number as her password for her email account, i.e., her password is $a_1a_2a_3a_4$, and each $a_i \in \{0, 1, \ldots, 9\}$. The attacker A who knows Alice's email account name simply tries to guess her password to log in. Suppose the email server did some basic protection that the online log in trials are limited to 6 times. Then what is the probability that A can successfully guess Alice's password and log in? (7 points)

B. Even though the guessing probability in above case is not negligible, A still wants to increase it and wants to learn Alice's password with 100% confidence. A breaches into the email server and notices that the password database entries are in the form of (name; H(pwd)), and A identifies the one corresponding to Alice easily. Describe step by step how A can fully reconstruct Alice's password. (8 points)

Q2: Biometric authentication is also one of the very popular ways of user authentication, for example, the fingerprint unlock of iPhone. List two advantages, and two disadvantages of biometrics over passwords. (5 points)

Q3: Other user authentication forms include something you have (e.g, a hardware token, ID card), or someone you known (e.g., when recover a social network account, to provide some information of a couple of your friends). List one advantage for each of them over the other. (5 points)

Q4: During the TLS handshake phase, the client has to authenticate the server, and share a key with the server to build a secure channel for future use, e.g., transmitting password. There are multiple possibilities for an attacker to weaken the security without breaking the key exchange, including man-in-the-middle attack and the replay attack.

1. Man-in-the-middle can happen as the attacker hijacks the communication channel between server and client, and then impersonates. For example, normally, Client sends g^a, Server sends g^b, then they both get g^{ab} as the shared key. But now, the attacker sends g^c to the client, claiming g^c is from the server, in this case, the client would consider g^{ac} as the shared key, which can be computed by the attacker. Briefly describe how TLS prevents such impersonation. (9 points)

2. In a replay attack, the attacker simply stores one session of the information he eavesdropped from the server, and sends in the future session to obtain advantage. Since now the messages were indeed from the server, thus cannot be considered as impersonation, but still, the message is not sent for the right time. Briefly describe how TLS handles such an issue. (8 points)

3. TLS/SSL enables two parties to build a secure channel, but still there are all kinds of security problems that may leak the server data, name two examples of security attacks that TLS/SSL cannot take care of. (8 points)

Part 3 (25 Points)

Q1: Many computer virus carries a "virus signature", and anti-virus software often uses known virus signatures for the purpose of detection.

- Briefly explain why virus signature exists (5 points)
- Briefly explain how a new virus may defeat a current version of anti-virus software, list at least two of them, and what we should do being an antivirus organization? (5 points).

Q2 Suppose you download and install an App to your phone, you see that it wants permission to "Send SMS messages" and to "Access your address-book". What threat might the App pose to your smartphone (and you), describe at least two? (5 points)

Q3**:** Consider a database table that includes a salary attribute. Suppose the three queries **sum, count**, and **max** (in that order) are made on the salary attribute/column, all conditioned on the same predicate/ condition involving other attributes (in SQL it is "Where xxx"). That is, a specific subset of records is selected based on the condition and the three queries are performed on that subset. Suppose that the first two queries (sum – means the summation of all the retrieved values, count – the number of retrieved items) are answered and the third query (max – the maximum among the retrieved values) is denied. Is there any extra information leaked beyond **sum, count**? You can give some potential examples as explanation. (10 points)