CR09 – Homework 2
Due date November 12, 11:59pm

The goal of this homework is to construct, step by step, a two-party protocol for secure biometric authentication. In this protocol, a user U holds a fingerprint F_U (modelled as a bitstring of length $n = 1024$) and wants to authenticate to a server S, who also holds the "authorized fingerprint" F_S (which is supposed to be the fingerprint of the user trying to authenticate). The server S wants to find out whether the user U is the authorized user, without revealing the authorized fingerprint; at the same time, the user does not want to reveal its local measured fingerprint. Because measurements of fingerprints can be noisy, the local measurement and the string stored on the server side might differ even for the authorized user. Hence, authentication is done by comparing the Hamming distance between F_U and F_S, denoted $\mathsf{HD}(F_U, F_S)$, to a threshold T (e.g. $T = 32$). Recall that for $x, y \in \{0,1\}^n$, we have $\mathsf{HD}(x, y) := \#\{i \in [n] | x_i \neq y_i\}$.

We recall that the Pedersen commitment over a group $\mathbb{G}$ of order p works as follows: a commitment to $m \in \mathbb{Z}_p$ with randomness $r \in \mathbb{Z}_p$ is of the form $c = g^m h^r$, where (g, h) is a pair of random generators of $\mathbb{G}$ which is assumed to be publicly available. Opening c amounts to revealing (m, r).

Remark: When a zero-knowledge proof is asked, it suffices to describe a Σ-protocol for the task (which can be converted to a zero-knowledge proof via the generic transform seen in class).

Question 1 (1 point). Assume for now that both U and S are honest. Using ElGamal encryption with the server's public key, describe a two-round protocol in which S speaks first and securely recovers $\mathsf{HD}(F_U, F_S)$.
Hint: Use that $\mathsf{HD}(x, y) = \mathsf{HW}(x) + \mathsf{HW}(y) - 2 \cdot \langle x, y \rangle$.

Question 2 (1 point). Describe an alternative, three-round protocol using ElGamal encryption with the user's public key, where U speaks first, and such that only S gets the result, while even an unbounded user does not learn anything about it.
Hint: the first flow is a bit-by-bit encryption of the user's input.

Question 3 (1 points). We now use the latter protocol (where messages are encrypted under U's public key but only S gets the end result) but we further assume that the user might behave dishonestly and learns whether authentication succeeded at the end of the protocol (the server still behaves honestly). Recall that authentication succeeds if $\mathsf{HD}(F_U, F_S) \leq T$, where F_S is the server's input, and F_U is the user's (possibly corrupted) input. We also assume that the fingerprint encoding is such that with high probability, the bitstring describing a fingerprint has roughly the same number of zeroes and ones.

Describe an attack that allows the user to recover any single bit of their choice from the authorized fingerprint, and to authenticate with probability $\approx 1/2$.

Question 4 (2 points). Describe a modification of the protocol, using zero-knowledge proofs, that circumvent the above attack. How much does this increase the communication complexity of the protocol? And its round complexity?

Question 5 (3 points). Assume now that the server might behave dishonestly as well, and that a bit-by-bit *Pedersen commitment* of the authorized fingerprint is publicly known. Describe a solution, using zero-knowledge proofs, where the server reveals an ElGamal encryption of the Hamming distance $\mathsf{HD}(F_U, F_S)$ and proves that it was computed honestly. How much does this increase the communication complexity of the protocol?
Note: several solutions exist. It is not necessary to find the best one.

Question 6 (2 points). Describe a zero-knowledge proof that allows the server to prove that the ElGamal encryption of the Hamming distance $\mathsf{HD}(F_U, F_S)$ encrypts a value below the threshold T, where T is some constant (e.g., 32 in our scenario).