TD 2 – Hash functions

Exercise 1. *Meet-in-the-middle preimage attack*

To build a compression function from a block cipher, an alternative to Davies-Meyer construction, known as PGV-13, is $f(h_{i-1}, m_i) = E(m_i, h_{i-1}) \oplus c$ where c is some fixed constant. It can be shown that in the ideal cipher model, a hash function H built from f using the Merkle-Damgård construction has *birthday-security* against both collision and preimage attacks.

The meet-in-the-middle attack allows to build a preimage for H, roughly in the same time needed to find a collision. We assume in the following that E is an ideal cipher block. *In this exercise, to simplify, we consider a variant of Merkle-Damgård construction without padding.*

1. **i.** Given h and t, what is the time needed to find m such that $f(h, m) = t$?
   ii. What can you say about the preimage security of f itself?
2. Adapt the proof of the birthday bound to prove the following: Let $y_1, \ldots, y_q, z_1, \ldots, z_q$ be uniformly and independently drawn from a size-N set, with $q \leq \sqrt{2N}$; Then the probability that there exist i and j such that $y_i = z_j$ is between $q^2/2N$ and q^2/N.
3. Back to the attack: Assume we sample q random elements $m_1, \ldots, m_q$ and compute $y_i = f(IV, m_i)$ for $i = 1$ to q, and similarly we sample q random elements $n_1, \ldots, n_q$ and compute $z_j = E^{-1}(n_j, t \oplus c)$ for all $j = 1$ to q. What is the probability that there exist i and j such that $y_i = z_j$?
4. Show how to build a two-block preimage of a given t, in time (roughly) $O(2^{n/2})$.

Exercise 2. *Multicollisions*

Let H be a hash function built from a compression function $f : \{0,1\}^n \times \{0,1\}^w \to \{0,1\}^n$ using the Merkle-Damgård construction. We assume $w > 2n$. For any c, let $H_c : \{0,1\}^{cw} \to \{0,1\}^n$ be a function built using the same Merkle-Damgård construction, but without padding and for messages of length cw exactly.

We are interested in finding *k-multicollisions*, that is a set of k messages $m_1, \ldots, m_k$ such that $H(m_1) = \cdots = H(m_k)$.

1. **i.** What time is needed to find a collision $f(IV, m_0^0) = f(IV, m_0^1)$?
   ii. Let $h_1 = f(IV, m_0^0) = f(IV, m_0^1)$. What time is needed to find a collision $f(h_1, m_1^0) = f(h_1, m_1^1)$?
   iii. Show how to find a 4-multicollision for H_2, and analyze the running time.
   iv. Explain how to turn this 4-multicollision for H_2 into a 4-multicollision for H.
2. Generalize the previous construction to build 2^t-multicollisions and analyze the cost.