TD 4 – Hash functions: The Kelsey-Schneier attack (2005)

Compression function. Let $f : \{0,1\}^n \times \{0,1\}^w \to \{0,1\}^n$ be a compression function, with $n \le w$, and let IV be some fixed initial value. For a message $\hat{m} = m_1 \| \cdots \| m_B$ of length $B \times w$, let $h_0 = IV$ and $h_i = f(h_{i-1}, m_i)$ for $i \ge 1$. Then we define F by $F(m_1 \| \cdots \| m_i) = h_i$ for all i. In particular, $F(m) = h_B$.
Merkle-Damgård construction. For a message $m \in \{0,1\}^*$, let $\text{pad}(m) = m \| 10 \cdots 0 \| \langle \text{length of } m \rangle$ be the padded version of m where the number of zeroes is adjusted to have $|\text{pad}(m)| = B \times w$ for some B. Then we define $H(m) = F(\text{pad}(m))$.
Kelsey & Schneier attack sketch. The idea of the second preimage attack of Kelsey & Schneier (2005) is to sample messages m_0 until $f(h_0, m_0) = h_i$ for some i. The expected number of samples before a success is $2^n/B$, assuming that $f(h_0, \cdot)$ behaves like a random function. Then, $\hat{m} = m_0 \| m_{i+1} \| \cdots \| m_B$ satisfies $F(\hat{m}) = H(m)$. Yet, since $\text{pad}(m)$ contains the length of m, there is no m' such that $\hat{m} = \text{pad}(m')$.
Birthday bound with two lists. Let $y_1, \ldots, y_q$ and $z_1, \ldots, z_q$ be uniformly and independently drawn from a size-N set. Then for $q \le \sqrt{N}$, $\frac{q^2}{2N} \le \Pr\left[\exists i, j, y_i = z_j\right] \le \frac{q^2}{N}$.

Exercise 1. *Expandable messages and second preimage attack*
An *expandable message of hash* h_{exp} is a set of messages $M_{\text{exp}} = \{m^1, m^2, \ldots\}$ such that $|m^i| = i \times w$ and $F(m^i) = h_{\text{exp}}$ for all i. Let M_{exp} be an expandable message of hash h_{exp}.

1. What is the cost of finding a one-block message m_0 such that $f(h_{\text{exp}}, m_0) = h_i$ for some i?
2. Explain how to produce a message m' such that $H(m') = H(m)$, given M_{exp} and m_0.
3. What is the cost of the attack, ignoring the cost of producing an expandable message? Why is this attack called a *long message second preimage attack*?

Exercise 2. *Expandable message from fixed points*
Let f be a compression function built from a block cipher E using Davies-Meyer construction: $f(h, m) = E_m(h) \oplus h$. We want to build an expandable message M_{exp} from a *fixed point for* f, that is from a pair (h_f, m_f) such that $f(h_f, m_f) = h_f$.

1. Let $m_f \in \{0,1\}^n$ be any one-block message, and $h_f = E_{m_f}^{-1}(0)$. Prove that (h_f, m_f) is a fixed point for f.
2. To build an expandable message M_{exp}, we adopt the following strategy: We produce a list of hashes $h = f(h_0, m_0)$ by sampling random blocks $m_0 \in \{0,1\}^n$; We produce a second list of hashes $h_f = E_{m_f}^{-1}(0)$ by sampling random blocks $m_f \in \{0,1\}^n$.
   i. Assume we found m_0 and (h_f, m_f) such that $f(h_0, m_0) = h_f$. Build an expandable message from this.
   ii. Prove that if we sample $2^{n/2}$ blocks m_0 and the same number of blocks m_f, the probability to get a collision is $\ge \frac{1}{2}$. *Assume that* $E(\cdot, 0)$ *and* $f(h_0, \cdot)$ *behave like random functions.*
3. Recap the steps of the full attack with this fixed point approach, and estimate its cost.

Exercise 3. *Expandable messages from multicollisions*
We are interested in finding *k-multicollisions* for F, that is a set of k messages $\hat{m}^1, \ldots, \hat{m}^k$ such that $F(\hat{m}^1) = \cdots = F(\hat{m}^k)$. If they all have distinct lengths, this is actually an expandable message.

1. i. Prove that for any ℓ_0, we can find $m_0 \in \{0,1\}^n$ and $m^0 \in \{0,1\}^{\ell_0 n}$ such that $f(h_0, m_0) = F(m^0)$, in time $O(2^{n/2})$. *We can fix the first* $(\ell_0 - 1)$ *blocks of* m^0.
   ii. Prove that once a collision $h_1 = f(h_0, m_0) = F(m^0)$ is found, we can in the same time find a collision $f(h_1, m_1) = F(m^1)$ where m^1 has ℓ_1 blocks.
   iii. Let $\ell_i = 1 + 2^i$ for all i and assume that we have found collisions $f(h_i, m_i) = F(m^i)$ for $i = 0$ to $t - 1$, where m^i has ℓ_i blocks. Prove that we can build a 2^t-multicollision for F, with messages of size tn to $(t + 2^t - 1)n$.
2. Recap the full attack with the multicollision and estimate its cost. *What condition must be satisfied by m to be able to find a second preimage?*