A collision-resistant hash function (CRHF) is a pair of algorithms \((\text{Gen}, \text{Hash})\) s.t.

\[
\begin{align*}
\text{Gen}(1^k) & \quad \text{outputs a key} \\
\text{Hash}_k : \{0,1\}^* & \rightarrow \{0,1\}^k \quad \text{where key} \leftarrow \text{Gen}(1^k)
\end{align*}
\]

and for ppt \(A \in \text{negl.}_k\) s.t.

\[
\Pr \left[ \text{key} \leftarrow \text{Gen}(1^k); (x_1, x_2) \leftarrow A(1^k, \text{key}) : x_1 \neq x_2 \text{ and } H_{\text{key}}(x_1) = H_{\text{key}}(x_2) \right] = \Omega(1)
\]

\Rightarrow \text{Applications: why is this a useful crypto primitive?}

1. **Digital signatures**: to sign a large doc \(D\)
   a) hash \(D\) down to \(k\) bits
   b) sign the hash of \(D\) using a digital signature scheme

2. **To efficiently compare two large docs** (to see if they are =)

**Aside**: why is it important that CRHF is a family of functions?

\(A:\) if alg is fixed, then \(A\) can be hardwired

Suppose that for every \(k\), there is just one \(H_{\text{key}_k}\), then let \(x_k, x_k'\) be s.t. \(H_{\text{key}_k}(x_k) = H_{\text{key}_k}(x_k')\) we know they exist
same as our prev def of CRHF except

\[ H_{key} : \{0,1\}^{2k} \rightarrow \{0,1\}^{k} \]

Suppose you have a CRHF that takes 2k-bit strings & hashes them down to k-bit strings. Can you use it to construct a CRHF family that is "unrestricted length"?

**TRY #1**

Input \( x = x_1 \ldots x_{2k} \)

but this is already a collision 😂 & repent until you get down to k-bits

**TRY #2**

Input \( x = x_1 \ldots x_{2k} \)

k-bits k-bits ≠ k-bits

\[
\begin{align*}
    y_1 &= H(x_1, x_2) \\
    y_i &= H(y_{i-1} \circ x_{i+1}) \quad \text{for } 2 \leq i \leq l-2 \\
    y_{l-1} &= H(y_{l-2} \circ \text{padded } x_k) \\
    y_l &= H(y_{l-1} \circ \text{len})
\end{align*}
\]

↑ \( x_{k} = 0^{l-2} \)

↑ length of \( x \) written in binary

\[ \text{Hash}(x) = \begin{cases} 
H(x_1 \circ x_2 \circ \ldots) & \text{if } 1x1 \leq k, \\
H(H(x_1 \circ x_2 \circ \ldots) \circ \text{len}) & \text{padded } x
\end{cases} \]
Hash key yields a CRHF family.

Pf. Suppose A outputs $x, x'$ s.t. $x \neq x'$ but $\text{Hash}_\text{key}(x) = \text{Hash}_\text{key}(x')$

Let $y_0, y_1, \ldots, y_i \text{ and } y_0', y_1', \ldots, y_i'$ be the intermediary results of computing $\text{Hash}_\text{key}(x)$ and $\text{Hash}_\text{key}(x')$.

Case 1: if $|x| \neq |x'|$ then $\text{len} \neq \text{len}'$ so we found a collision for $H$

$$y_i = H(y_{i-1} (\text{len})) = H(y_{i-1}' (\text{len}')) = y_i'$$

Case 2: $l = l'$ and $\forall i \in \mathbb{Z} y_i = y_i'$

but for some $i, x_i \neq x_i'$, then this is our collision.

Case 3: $l = l'$ and $\exists i \ s.t. y_i \neq y_i'$

Let $i$ be s.t. $y_i \neq y_i'$ but $\forall j < i \ y_j = y_j'$

$$y_i = H(y_{i-1} (x_{i+2})) = H(y_{i-1}' (x_{i+2}')) = y_i$$

**Remark:** Note construction & theorem generalize, we could have used $H_{\text{key}} : \{0,1\}^{k_1} \rightarrow \mathbb{Z}^{d \cdot k_2}$ where $k_1 > k_2$.

**Corollary:** To construct an unrestricted CRHF family, it is sufficient to construct a restricted length CRHF family.
Discrete log-based construction

- Group $G$ of prime order $q$, $q, h \in G$
  - $H_{q, h} : \{0, 1\}^* \rightarrow G$
  - $H_{q, h}(x) = q^{x_1} h^{x_2}$

- $X = x_1 \circledast x_2$ where $x_1, x_2 \in \mathbb{Z}_q$

Proof of collision resistance:
- Suppose $A$ outputs $(x_1, x_2)$ & $(x'_1, x'_2)$ such that $(x_1, x_2) \neq (x'_1, x'_2)$ but $q^{x_1} h^{x_2} = q^{x'_1} h^{x'_2}$

- $q > 2^k$ so every $k$-bit string corresponds to an integer $< q$

  - $q^{x_1-x'_1} = h^{x_1-x'_1}$
  - $h = q^{x_1-x'_1} (x_1-x'_1)$

- RSA modulus-based construction

Key Gen: RSA modulus $N$, $N = pq$ where $p, q$ are safe primes
- $p = 2p' + 1$
- $q = 2q' + 1$
- $\phi(N) = (p-1)(q-1) = 4p'q'$

- Let $Z \in \mathbb{QR}_N$ s.t. its order is $p'q'$
- key $= (N, Z)$

- $h_{NZ}(x) = Z^x \mod N$

- integer
IN PRACTICE, we use ad hoc hash functions.

SHA-1 by NIST 160 bits
broken by Google in 2016

high-level idea: from birthday paradox, only need \( \sqrt{160} \) trees to find collision

+ other crypto analysis to shave off a few more bits

Crypto analytic breakthrough 2004
(also broke MD5)

SHA-3 competition, Keccak 256, 512 bits