Definition: $\underline{\mathsf{Enc\text{-}Forge}_{\mathcal{A},\Pi}(n)}$: Run $\mathsf{Gen}(1^n)$ to obtain k. Adversary $\mathcal{A}$ is given input 1^n and access to an encryption oracle $\mathsf{Enc}_k(\cdot)$. They output ciphertext c. Let $m := \mathsf{Dec}_k(c)$, and let Q denote the set of all queries that $\mathcal{A}$ asked its encryption oracle. The output of the experiment is 1 iff $m \neq \perp$ and $m \notin Q$.

Definition: Unforgeable: A private-key encryption scheme Π such that for all PPT adversaries $\mathcal{A}$, $\Pr[\mathsf{Enc\text{-}Forge}_{\mathcal{A},\Pi}(n) = 1] \leq \mathtt{negl}(n)$.

Definition: Authenticated: A private-key encryption scheme that is CCA-secure and unforgeable.

Construction: Encrypt-and-authenticate: Given plaintext m, sender transmits $\langle c, t \rangle$, where $c \leftarrow \mathsf{Enc}_{k_E}(m)$ and $t \leftarrow \mathsf{Mac}_{k_M}(m)$. The receiver behaves as expected, obtaining m from $\mathsf{Dec}_{k_E}(c)$, and running $\mathsf{Vrfy}_{k_M}(m, t)$. It is likely the case here that t leaks information about the message (often, MACs are deterministic, breaking CPA-security), and so this is not an authenticated encryption scheme.

Construction: Authenticate-then-encrypt: Given plaintext m, sender transmits c, where $t \leftarrow \mathsf{Mac}_{k_M}(m)$ and $c \leftarrow \mathsf{Enc}_{k_E}(m||t)$. The receiver behaves as expected, decrypting $m||t$ from c, then checking $\mathsf{Vrfy}_{k_M}(m, t)$. If, for example, a CBC-mode-with-padding scheme is used, the decrypt algorithm will return a "bad padding" error, while if the padding passes, Vrfy will return an "authentication failure". This difference can leak information and allow for various attacks on the scheme, so this is not an authenticated encryption scheme.

Construction: Encrypt-then-authenticate: Given plaintext m, sender transmits $\langle c, t \rangle$, where $c \leftarrow \mathsf{Enc}_{k_E}(m)$ and $t \leftarrow \mathsf{Mac}_{k_M}(m)$. The receiver behaves as expected, checking $\mathsf{Vrfy}_{k_M}(c, t)$, then decrypting m as $\mathsf{Dec}_{k_E}(c)$. Of the three listed, this is the only one that is an authenticated encryption scheme (Assuming that Enc is CPA-secure, Mac is strongly secure, and k_E and k_M are chosen independently uniformly at random.)

There are 3 major types of network attacker attacks.

In a reordering attack, an attacker swaps the order of messages sent across a network, making c_2 arrive before c_1.

In a replay attack, an attacker resends messages later.

In a reflection attack, an attacker sends messages from a sender back to them at a later time, which the other person never sent.

The first two attacks can be prevented when A and B (the two people communicating across the network) keep counters, $\mathtt{ctr}_{A,B}$ and $\mathtt{ctr}_{B,A}$, of how many messages have been sent/received in each direction.

A reflection attack can either be prevented by having a reflection bit b to say who the sender is, or by having a different key-set for messages going different directions.

In the $\mathsf{Mac\text{-}forge}^{\text{1-time}}_{\mathcal{A},\Pi}$ experiment, adversary $\mathcal{A}$ outputs m', is given a tag $t' \leftarrow \mathsf{Mac}_k(m')$, then can calculate and think, then output (m, t), $m \neq m'$, which are verified as usual to determine success.

Definition: ε-secure (also one-time ε-secure): A MAC $\Pi = (\mathsf{Gen}, \mathsf{Mac}, \mathsf{Vrfy})$ such that for all (even unbounded) adversaries $\mathcal{A}$, $\Pr[\mathsf{Mac\text{-}forge}^{\text{1-time}}_{\mathcal{A},\Pi} = 1] \leq \varepsilon$.

Definition: Strongly universal: A function $h : \mathcal{K} \times \mathcal{K} \to \mathcal{T}$ such that for all distinct $m, m' \in \mathcal{M}$, and all $t, t' \in \mathcal{T}$, it holds that $\Pr[h_k(m) = t \wedge h_k(m1) = t'] = \frac{1}{|\mathcal{T}|^2}$, where the probability is taken over uniform choice of $k \in \mathcal{K}$.

Construction: : Let $h : \mathcal{K} \times \mathcal{M} \to \mathcal{T}$ be a strongly universal function. Define a MAC as follows: Gen: uniform $k \in \mathcal{K}$. Mac: given k, m, output tag $t := h_k(m)$. Vrfy: On input k, m, t, output 1 iff $t \stackrel{?}{=} h_k(m)$.

Theorem: : If h is a strongly universal function, then the above construction is a $\frac{1}{|\mathcal{T}|}$-secure MAC for messages in $\mathcal{M}$.

Theorem: : for any prime p, the function h defined as $h_{a,b}(m) = [a \cdot m + b \mod p]$, where $\mathcal{M} = \mathbb{Z}_p$, and $\mathcal{K} = \mathbb{Z}_p \times \mathbb{Z}_p$, so $(a, b) \in \mathcal{K}$, $m \in \mathcal{M}$, is strongly universal.

Definition: Hash function: A function with output length ℓ is a pair of PPT algorithms (Gen, H) such that $\mathsf{Gen}(1^n)$ outputs a key s, and H takes s and a string $x \in \{0,1\}^*$, and outputs a string $H^s(x) \in \{0,1\}^n$, assuming n is implicit in s.

Definition: Compression function (fixed-length hash function for inputs of length ℓ'): a hash function where H^s is only defined for inputs $x \in \{0,1\}^{\ell'(n)}$, and $\ell'(n) > \ell(n)$.

Definition: $\underline{\mathsf{Hash\text{-}Coll}_{\mathcal{A},\Pi}(n)}$: $s \leftarrow \mathsf{Gen}(1^n)$. Adversary $\mathcal{A}$is given s and outputs x, x'. (If Π is fixed-length, then $x, x' \in \{0,1\}^{\ell'(n)}$.) The output is 1 (success) iff $x \neq x'$ but $H^s(x) = H^s(x')$.

Definition: Collision resistant: A has function $\Pi = (\mathsf{Gen}, H)$ such that for all PPT adversaries $\mathcal{A}$, $\Pr[\mathsf{Hash\text{-}Coll}_{\mathcal{A},\Pi}(n) = 1] \leq \mathtt{negl}(n)$.

Definition: Second-preimage resistance (target-collision resistance): A hash function such that given s and x, an adversary cannot find x' such that $x' \neq x$ and $H^s(x) \neq H^s(x')$.

Definition: Preimage resistance: A hash function such that given s and y, an adversary cannot find x such that $H^s(x) = y$.

Construction: Merkle-Damgård: Let (Gen, h) be a fixed-length hash function for inputs of length $2n$ and with output length n. Construct (Gen, H) as follows: $\mathsf{Gen} = \mathsf{Gen}$, H: given s and $x \in \{0,1\}^*$ of length $L < 2^n$, let $B = \lceil \frac{L}{n} \rceil$, pad x so its length is a multiple of n. Consider the padded result as n-bit blocks $x_1, \ldots, x_B$. Set $x_{B+1} = L$. Set $z_0 = 0^n$, as the IV. For $i = 1, \ldots, B+1$, let $z_i = h^s(z_{i-1}||x_i)$. Output z_{B+1}.

Theorem: If (Gen, h) is collision resistant, then so is (Gen, H).

Construction: Hash-and-MAC: Let $\Pi = (\mathsf{Mac}, \mathsf{Vrfy})$ be a MAC for length $\ell(n)$, let $\Pi_H(\mathsf{Gen}_H, H)$ be a hash function, with output length $\ell(n)$. Construct MAC $\Pi' = (\mathsf{Gen}', \mathsf{Mac}', \mathsf{Vrfy}')$ as follows: Gen': Takes 1^n, choses uniform $k \in \{0,1\}^n$, $s \leftarrow \mathsf{Gen}_H(1^n)$, outputs key $k' = \langle k, s\rangle$. Mac': Given $\langle k, s\rangle$, $m \in \{0,1\}^*$, output $t \leftarrow \mathsf{Mac}_k(H^s(m))$. Vrfy': Given $\langle k, s\rangle$, $m \in \{0,1\}^*$, tag t, output 1 iff $\mathsf{Vrfy}_k(H^s(m), t) = 1$.

Theorem: : If Π is a secure MAC and Π_H is collision resistant, the above construction is a secure MAC for arbitrary-length messages.

Construction: HMAC: Let (Gen_H, H) be a Merkle-Damgård-generated hash function on (Gen_H, h) taking inputs of length $n + n'$. Let $\mathtt{opad}$and $\mathtt{ipad}$be fixed constants of length n'. Define a MAC as follows: Gen: Given 1^n, $s \leftarrow \mathsf{Gen}_H(1^n)$, uniform random $k \in \{0,1\}^{n'}$. Output key $\langle s, k\rangle$. Mac: Given $\langle s, k\rangle$ and $m \in \{0,1\}^*$, output $t := H^s\left((k \oplus \mathtt{opad})||H^s((k \oplus \mathtt{ipad})||m)\right)$. Vrfy: Given $\langle s, k\rangle$, $m \in \{0,1\}^*$, tag t, output 1 iff t recomputes correctly.

Definition: Weakly collision resistant: A Hash function (Gen_H, H) defined as a Merkle-Damgård transform, except with $k = IV$ being uniformly chosen from $\{0,1\}^n$, such that every PPT adversary $\mathcal{A}$ has at most negligible success finding a collision (without knowing k.).

Theorem: Let $k_{out} = h^s(IV||(k \oplus \mathtt{opad}))$, $\hat{y}$ be the length-padded y, including anything before it, $\widetilde{\mathsf{Mac}}_k(y) = h^s(k||\hat{y})$, and $G^s(k) = h^s(IV||(k \oplus \mathtt{opad}))||h^s(IV||(k \oplus \mathtt{ipad})) = k_{out}||k_{in}$. If G^s is a PRG for any s, $\widetilde{\mathsf{Mac}}_k(y)$ is a secure fixed-length mac for messages of length n, and (Gen_H, H) is weakly collision resistant, then HMAC is a secure MAC for arbitrary-length messages.

Definition: Birthday problem/attack: Out of n distinct "days", if $\sqrt{n}$ "people" are chosen, there is a 50% chance that two of them will share a birthday. This places a lower bound of $2\log(T)$ bits on the size of a hash function, where T is the time we want to run the collision-attack in.

Construction: Birthday Attack: (small space). Start with random valid input x_0, then repeatedly compute $x_i = H(x_{i-1})$ and $x_{2i} = H(H(x_{2(i-1)}))$. If they are ever equal, collision has occurred in $x_0, \ldots, x_{2(i-1)}$. Calculate each x_j, x_{j+i}, and we will find a collision. This runs in $\Theta(2\ell/2)$ time, where ℓ is the length of the output.

Definition: Random Oracle Model: Model in which the oracle O chooses its function H at random when instantiated, and so probabilities are also taken over the choice of the function H. It is then used in whichever function needed. The values of H are considered to be computed the first time they are requested.

Construction: If $\ell_{out} > \ell_{in}$, a random oracle can be used as a pseudorandom generator. If $\ell_{out} < \ell_{in}$, a random oracle is collision resistant. If $F_k(x) = H(k||x)$, where $|k| = |x| = n$, $\ell_{out} = n$, $\ell_{in} = 2n$, then F is a pseudorandom function, where H is the pseudorandom oracle.

In general, a proof of security in the random oracle model is significantly better than no proof at all. There have been no successful real-world attacks on schemes proven secure in the random-oracle model, when the random oracle was instantiated properly. Most cryptographic hash functions should not be used "off the shelf" to instantiate a random oracle model.

Definition: Virus Fingerprinting: Process by which a virus scanner stores hashes of known viruses, and compares hashes of email attachments and newly downloaded programs to these known viruses.

Definition: Deduplication: Process of comparing hashes of new files to hashes of already stored files to eliminate the storing of duplicates. Especially used in the context of cloud storage among multiple users.

Definition: P2P file-sharing: Processing of storing hashes of available files, allowing for easy requests, etc.

Definition: Merkle-Damgård Tree: A tree constructed from 2^t by placing a file at each leaf of a t-level tree, then computing the hash of each pair of files, then each pair of hashes, and so on, until a single root hash is computed. This hash is then stored. Often denoted $\mathcal{MT}_t$.

Theorem: Let (Gen_H, H) be collision resistant. Then $(\mathsf{Gen}_H, \mathcal{MT}_t)$ is also collision resistant for any fixed t.

Construction: Password Hashing: On a computer, the hash of a password, hpw, will be stored, and when the user enters their password, $H(pass) \stackrel{?}{=} hpw$, if so, authenticated. To prevent dictionary attacks, sometimes a salt is used to calculate $H(s, pass)$.

Definition: min-entropy: A probability distribution $\mathcal{X}$ has m bits of min-entropy if for every fixed value x it holds that $\Pr_{X \leftarrow \mathcal{X}}[X = x] \leq 2^{-m}$. That is, even the most likely outcome occurs with probability at most 2^{-m}.

Definition: LFSR: Linear Feedback Shift Register. Very efficient to implement in hardware. Consists of an array of n registers, $s_{n-1}, \ldots, s_0$ with n feedback coefficients, $c_{n-1}, \ldots, c_0$. The size of the array is called the degree of the LFSR. On each clock tick, s_0 is the output bit, all bits are shifted right by 1 register, and s_{n-1} is set to the XOR of some subset of the other registers, defined as those where $c_i = 1$.

These are insecure because the initial state, feedback coefficients, and all future bits, can be determined from watching at most $2n$ consecutive bits of output. We can improve the security by adding non-linear combinations to compute s_{n-1}, and it is possible to define such functions with good statistical properties. Trivium is one such stream cipher.

Definition: RC4: a similar algorithm that also involves bit swaps, is used in many security situations, but is known to have vulnerabilities.

Definition: Avalanche Effect: In any block cipher, a "small change" to the input must affect every bit of the output.

Definition: Feistel Network: A network which operates in a series of rounds. In each round, a keyed round function is applied. This function need not be invertible. In a balanced Feistel Network, the ith round of function $\hat{f}_i$ takes as input a sub-key x_i and an $\ell/2$-bit string and outputs an $\ell/2$-bit string.

Construction: Feistel Network: For round i of a Feistel network, divide the input into two halves, L_{i-1} and R_{i-1}, each of length $\ell/2$, where ℓ is the block length of the cipher. The output (L_i, R_i) is defined as $L_i := R_{i-1}, R_1 = L_{i-1} \oplus f_i(R_{i-1})$. In an r-round Feistel network, the ℓ-bit input becomes (L_0, R_0), and the output is the ℓ-bit value (L_r, R_r).

Theorem: Let F be a keyed function defined by a Feistel Network. Then regardless of the round functions $\{\hat{f}_i\}$, and the number of rounds, F_k is an efficiently invertible permutation for all k.

Definition: DES: Data Encryption Standard. Originally a 16-round Feistel Network with a block length of 64 bits and a key length of 56 bits. Vulnerable to brute-force attacks, but the strengthened triple-DES is widely used today.

Construction: DES $\hat{f}$. $\hat{f}(k_i, R)$, with $k_i \in \{0,1\}^{28}$, $R \in \{0,1\}^{32}$, R is expanded to 48-bit R' by duplicating the first half of the bits, $R' \oplus k_i$ is computed, split and passed through an S-box which takes 6-bit inputs to 4-bit outputs, and these 4-bit outputs are mixed to produce the 32-bit output. DES uses 16 rounds.

The S-boxes were carefully designed to be 4-to-1 functions, and changing any 1 bit of the input changes at least 2 bits of the output. The mixing was designed that the output from any S-box affects the input to six of the S-boxes in the next round. Therefore, the mangle function exhibits a strong avalanche effect, which will, after 8 rounds, affect all 64 bits of output. Since DES uses 16 rounds total, similar inputs yields independent-looking outputs.

After 30 years, the best known practical attack on DES is still an exhaustive search through its key space. The 56-bit key length is such that such an attack *is* feasible.

Definition: Triple Encryption: Using 3 keys, $F'_{k_1,k_2,k_3}(x) := F_{k_3}(F^{-1}_{k_2}(F_{k_1}(x)))$. Using only 2, $F'_{k_1,k_2}(x) := F_{k_1}(F^{-1}_{k_2}(F_{k_1}(x)))$. Triple-DES is constructed using DES with either of these variants.

Definition: AES: Advanced Encryption Standard. Has a state array, which is a 4x4 array. Initially the input of the cipher. Then, consists of 4 stages. Stage 1: Add Round Key: A 128-bit sub-key is derived from the master key, and interpreted as a 4x4 array of bytes. The state array is XORed with the sub-key. Stage 2: Sub Bytes: The state array is mixed at the byte level according to a fixed lookup table S. Stage 3: Shift Rows: The bytes in each row are shifted to the left by 0, 1, 2, and 3 places respectively, from the top. Stage 4: Mix Columns: An invertible transformation is applied to each column. It has the property that if the inputs differ in $b > 0$ bytes, the outputs differ in at least $5 - b$ bytes. In the final round, Mix Columns is replaced with AddRoundKey. To date, there have been no practical attacks significantly better than an exhaustive key-search, so AES is an excellent choice for any cryptographic scheme that requires a (strong) pseudorandom permutation.

Definition: Ideal Cipher Model: A strengthening of the random-oracle model in which all parties have access to an oracle for a random keyed permutation $F : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^\ell$ and F^{-1}.

Construction: Davies-Meyer: Let F be a block cipher with n-bit key length and ℓ-bit block length. The compression function is $h : \{0,1\}^{n+\ell} \to \{0,1\}^\ell$ by $h(k, x) := F_k(x) \oplus x$.

Theorem: If F is an ideal cipher, then Davies-Meyer yields a collision-resistant compression function. From the HW, $F_k(x) \oplus x \oplus k$ also does, but $F_k(x)$ and $F_k(x) \oplus k$ do not.

Care must be taken when instantiating Davies-Meyer with a particular block cipher, for example, DES causes issues.

MD5 is bad and should not be used, SHA-0 is flawed, SHA-1 has known slight flaws that make it easier to theoretically crack, but has not produced any collisions. It is not recommended for use, SHA-2 seems to be secure, and can be used. SHA-3 is rather powerful, and unusual, and considered very, very secure.