Cryptography notes
CS 6260 (Fall 2012) and CS 7560 (Spring 2013)

Christopher Martin
chris.martin@gatech.edu

1 Symmetric cryptography scheme

Key space	$\mathcal{K}$
Message space	$\mathcal{M}$
Cypher space	$\mathcal{C}$
Key generator	$\mathsf{Gen} : \phi \to \mathcal{K}$
Encryption function	$\mathsf{Enc} : \{\mathcal{K} \times \mathcal{M}\} \to \mathcal{C}$
Decryption function	$\mathsf{Dec} : \{\mathcal{K} \times \mathcal{C}\} \to \mathcal{M}$

2 Information theoretic security

Information theoretic security repels even resource-unbounded attackers. Shannon secrecy and perfect secrecy are equivalent definitions of information theoretic security for symmetric cryptography schemes.

Information security turns out to be a stronger definition than necessary, so generally we will instead consider computational security, which will impose computation bounds on attackers.

Shannon secrecy A scheme is Shannon-secret iff the ciphertext reveals no additional information about the message.

$\forall$ distribution D over M, $\bar{m} \in \mathcal{M}$, $\bar{c} \in \mathcal{C}$:

$$\Pr_{\substack{k \leftarrow \mathsf{Gen} \\ m \in D}} \left[\, m = \bar{m} \,|\, \mathsf{Enc}_k(m) = \bar{c} \,\right] = \Pr_{m \in D} \left[\, m = \bar{m} \,\right]$$

Perfect secrecy A scheme is perfectly secret iff the distributions of ciphertexts for any two messages are identical.

$\forall\, m_1, m_2 \in \mathcal{M},\; c \in \mathcal{C}$:

$$\Pr_{k_1 \leftarrow \mathsf{Gen}}\left[\, \mathsf{Enc}_{k_1}(m_1) = c \,\right] = \Pr_{k_2 \leftarrow \mathsf{Gen}}\left[\, \mathsf{Enc}_{k_2}(m_2) = c \,\right]$$

This model considers only a single message and ciphertext, so although a one-time pad is perfectly secret, a "two-time pad" is not.

Theorem 1. Perfect secrecy $\Rightarrow |\mathcal{K}| \geq |\mathcal{M}|$.

Proof. Let D be uniform over $\mathcal{M}$, and let $\bar{m} \in \mathcal{M}$, $\bar{c} \in \mathsf{Enc}_k(\bar{m})$, $\bar{k} \in \mathcal{K}$ be arbitrary. Let $S = \{\mathsf{Dec}_k(\bar{c}) : k \in \mathcal{K}\}$. $|S| < |\mathcal{M}|$ because Dec must be deterministic.

$$\Pr_{m \leftarrow D}[\; m = \hat{m} \;] = \frac{1}{|\mathcal{M}|}\ \forall\, \hat{m} \in \mathcal{M}\ .$$

If we choose $\hat{m} \in \mathcal{M} \backslash S$, then

$$\Pr_{\substack{m \leftarrow D,\\ k \leftarrow \mathsf{Gen}}}[\; m = \hat{m} \,|\, \mathsf{Enc}_k(m) = \bar{c} \;] = 0\ .$$

$\square$

3 One-way functions

A function $f : \{\mathtt{0}, \mathtt{1}\}^* \to \{\mathtt{0}, \mathtt{1}\}^*$ is a one-way-function if:

Easy to compute: $\exists$ a *ppt* algorithm F such that $F(x) = f(x)\ \forall\, x$.

Hard to invert: $\forall$ *nuppt* A, the advantage

$$\Pr_{x \leftarrow \{\mathtt{0},\mathtt{1}\}^n}[\; A(\underbrace{\mathtt{1}^n}_{\text{helps bound } A}, f(x)) \in \underbrace{f^{-1}(f(x))}_{\text{preimages of } f(x)} \;] = \mathrm{negl}(n)$$

The domain and range are given as $\{\mathtt{0}, \mathtt{1}\}^*$ for convenience, but any $D_n \to R_n$ is okay if D_n can be efficiently sampled.

Subset-sum f_{ss}: $(\mathbb{Z}_N)^n \times \{\mathtt{0}, \mathtt{1}\}^n \to (\mathbb{Z}_N)^n \times \mathbb{Z}_N$, $N = 2^n$
$f_{\mathrm{ss}}(a_1, \ldots, a_n, b_1, \ldots, b_n) = (a_1, \ldots, a_n, S = \sum_{i=1}^{n} a_i b_i \bmod N)$
Subset-sum is NP-complete. This might be a one-way function.

Prime multiplication $f : (\mathrm{primes}_n)^2 \to \mathbb{N}$
$f(x, y) = xy$ might be a one-way function.

Weak one-way functions For a weak one-way function, any adversary has advantage $\leq 1 - \delta$ for some $\delta = \frac{1}{\text{poly}(n)}$. Multiplication might be a weak one-way function, because there are a lot of primes.

Hardness amplification We can use a weak one-way function to find a one-way function. If f is δ-OWF, then f' is strong-OWF where

$$f'(x_1, \ldots, x_m) = f(x_1), \ldots, f(x_m)$$

Assume $\mathcal{A}'$ with non-negl advantange $\alpha(n)$ against f'.

$$G_i \equiv \left\{ x_i : \Pr_{\substack{x_j \leftarrow \{\mathtt{0,1}\}^n, \\ j \neq i}} [\ \mathcal{A}'(f'(x' = (x_1 \ldots x_m))) \text{ inverts }] \geq \frac{\alpha}{2m} \right\}$$

For $m \geq \frac{2n}{\delta}$, $\exists\ i$ such that $\frac{|G_i|}{2^n} \geq 1 - \delta/2$. Proof: Suppose not.

$$\begin{aligned}
\Pr_{x'}[\ \mathcal{A}'(f'(x')) \text{ inverts }] &\leq \Pr_{x'}[\ \mathcal{A}' \text{ inverts } \wedge \text{ every } x_i \in G\ ] \\
&\quad + \sum_{i=1}^{m} \Pr[\ \mathcal{A}' \text{ succeeds } \wedge x_i \notin G\ ] \\
&\leq \Pr_{x}[\text{ every } x_i \in G_i\ ] + \sum_{i=1}^{m} \Pr[\ \mathcal{A}' \text{ succeeds} \mid x_i \notin G\ ] \\
&\leq (1 - \frac{\delta}{2})^{2n/\delta} + \sum_{i=1}^{m} \frac{\alpha}{2m} \\
&\leq 2^{-n} + \frac{\delta}{2} < \delta
\end{aligned}$$

Construct $\mathcal{A}$ against f: $\mathcal{A}(\mathtt{1}^n, y = f(x_i)) \equiv$ Repeat: Choose all of the x_j for $j \neq i$, invoke $\mathcal{A}'(f'(x')))$, win if $\mathcal{A}'$ succeeds.

$$\Pr_{x_i}[\ \mathcal{A} \text{ inverts }] \approx \Pr_{x_i}[x_i \text{ is good}] \geq 1 - \delta/2$$

Family of one-way functions $F = \{f_s : D_s \rightarrow R_s\}$

Easy to sample a function: $\exists$ *ppt* $S(\mathtt{1}^n)$ outputs some $f_s \in F$

Easy to sample from D_s

Easy to evaluate $f_s(x)$

Hard to invert: $\forall$ *nuppt* $\mathcal{A}$,

$$\text{Adv}_F(\mathcal{A}) = \Pr_{\substack{s \leftarrow S(1^n), \\ x \leftarrow D_s}} [\, \mathcal{A}(1^n, s, f_s(x)) \in f_s^{-1}(f_s(x)) \,] = \text{negl}(n)$$

Family of subset-sum functions

$$F_{\text{ss}} = \left\{ f_{\vec{a}}(b_1, \ldots, b_n) = \sum_i a_i \cdot b_i \bmod N \right\}$$

Rabin's function $f_N : \mathbb{Z}_N^* \to \mathbb{Z}_N^*$, $N = pq$ for distinct primes p, q, $f_N(x) = x^2 \bmod N$

Rabin OWF family $\iff$ factoring is hard Chinese remainder theorem: $h : \mathbb{Z}_N \to \mathbb{Z}_p \times \mathbb{Z}_q$
p, q distinct primes, $N = pq$
$h(x) = (x \bmod p, x \bmod q)$
Ring homomorphism: $h(x \cdot y) = h(x) \cdot h(y)$, $h(x + y) = h(x) + h(y)$
... is an isomorphism (a bijection)

$\left|\mathbb{QR}_p^*\right| = (p-1)/2$ for any odd prime p: because $a^2 = (-a)^2 = (p-a)^2 \bmod p$. $\mathbb{Z}_p$ is a field so there are ≤ 2 square roots for any element. Therefore $x \to x^2 \mod p$ is a 2-to-1 function $\mathbb{Z}_p^* \to \mathbb{QR}_p^*$

$\mathbb{Z}_N^* \cong \mathbb{Z}_p^* \times \mathbb{Z}_q^*$, so $\mathbb{QR}_N^* \cong \mathbb{QR}_p^* \times \mathbb{QR}_q^* \implies |QR_N^*| = \frac{p-1}{2}\frac{q-1}{2} = \frac{\phi(N)}{4}$
For $(a_p, a_q) \in \mathbb{QR}_p^* \times \mathbb{QR}_q^*$, $(\pm\sqrt{a_p}, \pm\sqrt{a_q})^2 = (a_p, a_q)$, so Rabin is 4-to-1.

Lemma: Given any $x_1, x_2 \in \mathbb{Z}_N^*$ such that $x_1^2 = x_2^2$ but $x_1 \neq \pm x_2 \bmod N$, the factorization of N can be computed efficiently. Proof: $x_1^2 = x_2^2 \bmod p$ and $\bmod q$. So $x_1 = \pm x_2 \mod p$ and $x_1 = \pm x_2 \mod q$, but the $\pm$ for these two statements cannot both be $+$ or both $-$. Otherwise, by Chinese remainder theorem, $x_1 = x_2$. w.l.o.g. $x_1 = +x_2 \bmod p$ and $x_1 = -x_2 \bmod q$. So $p \mid (x_1 - x^2)$ and $q \nmid (x_1 - x_2)$, and $\gcd(N, x_1 - x_2) = p$.

If factoring is easy: $\mathcal{A}(N = pq, y = x^2 \mod N) \equiv$ Factor N into p, q. Compute $\sqrt{y \bmod p} \in \mathbb{Z}_p^*$, $\sqrt{y \bmod q} \in \mathbb{Z}_q^*$ (proof omitted). Reconstruct $\sqrt{y} \in \mathbb{Z}_N^*$ using Chinese remainder theorem bijection (proof omitted).

If Rabin inversion is easy: Assume $\exists\, \mathcal{B}$ for which $\Pr_{N \leftarrow pq, X \leftarrow \mathbb{Z}_N^*}[\mathcal{B}(N, y = x^2 \bmod N) \in \sqrt{y}] \geq \alpha(n)$ non-negl(n). Build $\mathcal{A}(N = pq)$: Choose $x \leftarrow \mathbb{Z}_N^*$, let $y = x^2 \mod N$. $\mathcal{B}(N, y) = x' \in \mathbb{Z}_N^*$ such that $(x')^2 = y \mod N$ (w.p. $\geq \alpha(n)$). If $x \neq \pm x' \mod N$ (w.p. $1/2$), we can use the lemma. Overall, succeeds w.p. $\geq \alpha(n)/2$.

4 Psuedo-random generators (PRGs)

If $\exists$ a PRG G with output length $n+1$, $\exists$ a PRG G_ℓ with output length $\ell(n) = \text{poly}(n)$:

$$G_\ell(x) = (h(x), h(f(x)), h(f^2(x)), h(f^3(x)), \ldots, h(f^{\ell-1}(x)))$$

4.1 PRG implies PRF

PRG with $\ell(n) = 2n$ implies a PRF family. Write $G(x) = G_0(x) \,\|\, G_1(x)$ and:

$$f_s(x) = G_{x_n}(\ldots G_{x_2}(G_{x_1}(s))) \ldots)$$

Given $\mathcal{A}$ that distinguishes f, build $\mathcal{B}(z = z_0 \,\|\, z_1)$: When $\mathcal{A}$ queries with x, answer with $G_{x_n}(\ldots G_{x_2}(z_{x_1}) \ldots)$. If $z = G(s)$, then $\mathcal{B}$ simulates f_s. If $z \leftarrow \mathcal{U}_{2n}$, $\mathcal{B}$'s responses are computationally indistinguishable from random. TODO: Why isn't this argument that simple?

5 Pseudo-random functions

Uniformly random function U is a random variable chosen uniformly from the set of all functions $\{0,1\}^m \to \{0,1\}^n$.

Pseudo-random function A PRF belongs to a family of functions $F : \{0,1\}^\ell \times \{0,1\}^m \to \{0,1\}^n$. Write $F_k(\cdot)$ to denote $F(k, \cdot)$.

Distinguishing advantage Consider an adversary $\mathcal{A}$ who knows F, having oracle access to F_k where k was chosen uniformly at random, trying to distinguish the oracle's responses from a random function. The distinguishing advantange of $\mathcal{A}$ against F is

$$\text{Adv}_F^{\text{prf}}(\mathcal{A}) \equiv \Pr_{k \in \{0,1\}^\ell}\left[\, \mathcal{A}^{F_k} \text{ accepts} \,\right] - \Pr_U\left[\, \mathcal{A}^U \text{ accepts} \,\right] .$$

In time $O(t)$, we can brute-force t keys to get advantage $t/2^\ell$.

(t,q)-bounded adversary $\begin{array}{r|l} t & \text{Running time} \\ q & \text{Number of queries} \end{array}$

(t,q,ε)-secure PRF F is (t,q,ε)-secure iff $\forall$ (t,q)-bounded $\mathcal{A}$,

$$\text{Adv}_F^{\text{prf}}(\mathcal{A}) \leq \varepsilon .$$

Examples of reasonable constants

t	2^{128}
q	2^{64} or 2^{32}
ε	2^{-128}

Existence The existence of secure PRFs has not been proven, but there are some functions that have never been broken and are widely assumed to be PRFs.

6 Goldreich-Levin

Given $\overline{F}(x)$ OWF, define $F(x,r) \equiv (\overline{F}(x), r)$ for $|x| = |r|$. Then

$$h(x,r) \equiv \langle x, r\rangle \bmod 2 = \sum_{j=1}^{n} x_j \cdot r_j \bmod 2$$

is hard-core for F.

Suppose $\exists\, \mathcal{A}$ and non-negl δ s.t.

$$\Pr_{x,r \leftarrow \{\mathtt{0},\mathtt{1}\}^n}[\ \mathcal{A}(\overline{F}(x), r) = \langle x, r\rangle\ ] \geq \frac{1}{2} + 2\delta$$

Then we'll construct $\mathcal{B}$ s.t.

$$\Pr_{x \leftarrow \{\mathtt{0},\mathtt{1}\}^n}[\ \mathcal{B}(\overline{F}(x)) \in \overline{F}^{-1}(\overline{F}(x))\ ] \geq \delta$$

Let $\mathcal{O}(\cdot) \equiv \mathcal{A}(y, \cdot)$.

Define SC(z, $\{r\}_m \in \{\mathtt{0},\mathtt{1}\}^n$, $\{b\}_m \in \{\mathtt{0},\mathtt{1}\}$) for $m = O(\log n)$. Assume that $b_j = \langle x, r_j\rangle$ (we will end up trying all the $\{b\}$).

For each subset $S \subseteq [m]$: Let $b_S = \sum_{j\in S} b_j$ and $r_s = \sum_{j \in S} r_j$. So $\langle x, r_s\rangle = \left\langle x, \sum_{j\in S} r_j \right\rangle = \sum_{j\in S} \langle x, r_j\rangle = \sum_{j \in S} b_j = b_S$. Then $c_S \leftarrow \mathcal{O}(z + r_s) - b_S$. Return the majority value of c_S.

$\mathcal{B}$:

For each $j \in [m]$ choose $r_j \leftarrow \{\mathtt{0},\mathtt{1}\}^n$ randomly.

For each $\{b\} \in \{\mathtt{0},\mathtt{1}\}^m$:

For each $k \in [n]$ let $x'_k \leftarrow \mathrm{SC}(e_k, \{r\}_m, \{b\}_m)$

Let $x' = x'_1 \ldots x'_n$. Output x' if it is an inversion.

Over random choice of the r_j, the r_S are uniformly random and pairwise independent for all $S \neq \phi$.

$X_S \equiv \mathrm{ind}(\mathcal{O}(z + r_s)$ is correct$(= \langle x, z + r_s\rangle))$

We're interested in whether $X = \sum_{S \subseteq [m]} X_S > \frac{m}{2}$

Local list-decoding the Hadamard code With a weak SC and $< \frac{1}{4}$ fraction of errors, can recover the unique x. GL says with strong SC at $\frac{1}{2} - \delta$ fraction of errors, we can recover all possible x.

7 Reduction

Karp (many-to-one) reduction Reduction from A to B transforms an instance of A to an instance of B.

Cook (Turing) reduction Reduction from A to B solves A using a subroutine that solves B.

Key recovery security F is (t, q, ε)-kr-secure iff $\forall$ (t, q)-bounded $\mathcal{A}$,

$$\mathrm{Adv}_F^{\mathrm{kr}} \equiv \Pr_{k \in \{0,1\}^\ell} \left[\mathcal{A}^{F_k(\cdot)} \text{ outputs } k \right] \leq \varepsilon\ .$$

Theorem 2. If F is a (t, q, ε)-secure PRF for $q < 2^m$, then F is (t', q', ε')-kr-secure for $t' \approx t$, $q' = q - 1$, $\varepsilon' = \epsilon + 2^{-n}$.

Proof. Cook reduction. For any kr-adversary $\mathcal{A}'$ running in time t' and making $q' < 2^m$ queries, let $\mathcal{A}$ be the PRF adversary:

$k' \leftarrow \mathcal{A}'(\mathcal{O})$
$x \leftarrow$ a value that $\mathcal{A}'$ did not query with
$y \leftarrow \mathcal{O}(x)$
Accept iff $y = F_{k'}(x)$

$\mathcal{A}$ runs in time $t \approx t'$ and makes $q = q' + 1$ queries.

$$\mathrm{Adv}_F^{\mathrm{prf}}(\mathcal{A}) \geq \mathrm{Adv}_F^{\mathrm{kr}}(\mathcal{A}') - 2^{-n}\ . \qquad \square$$

Example PRF construction For PRF $F : \{\mathtt{0},\mathtt{1}\}^\ell \times \{\mathtt{0},\mathtt{1}\}^n \to \{\mathtt{0},\mathtt{1}\}^n$, $F'_k(x) \equiv F_k(F_k(x)) \| F_k(\overline{F_k(x)})$

Theorem 3. F' is $(t, \dfrac{q}{3}, \varepsilon + \dfrac{q^2}{2^n})$-secure.

Proof. Let $\mathcal{A}'$ be an attacker on F'. Define $\mathcal{A}$ as:

$\mathcal{O}' \equiv \mathcal{O}(\mathcal{O}(x)) \| \mathcal{O}(\overline{\mathcal{O}(x)})$ (done with 3 queries to $\mathcal{O}$)
Accept iff ${\mathcal{A}'}^{\mathcal{O}'}$ accepts

$\mathcal{O}'(x)$ simulates F' perfectly, so $\Pr_k\left[\mathcal{A}^{F_k}\right] = \Pr_k\left[\mathcal{A}'^{F'_k}\right]$.

$\mathcal{O}'$ does not simulate U perfectly, but it is close. We have independence as long as all of the $\mathcal{O}(x)$, $\overline{\mathcal{O}(x)}$ are distinct. Using union bound, this probability $\leq \frac{q^2}{2^n}$ $\square$

8 Pseudo-random permutations

In a permutation family $F : \{0,1\}^\ell \times \{0,1\}^n \to \{0,1\}^n$, every F_k is bijective.

A secure PRP is computationally indistinguishable from a uniformly random permutation.

Some well-known PRPs

DES	$\ell = 56$	$n = 64$
AES$_{128}$	$\ell = 128$	$n = 128$
AES$_{192}$	$\ell = 192$	$n = 128$

Strong PRP / block cipher Attackers with oracle access to both F and F^{-1} have small advantage.

$$\mathrm{Adv}_F^{\mathrm{sprp}} \equiv \Pr_k\left[\mathcal{A}^{F_k, F_k^{-1}} \text{ accepts}\right] - \Pr_P\left[\mathcal{A}^{P, P^{-1}} \text{ accepts}\right] \leq \varepsilon$$

PRF/PRP switching lemma If G is a (t, q, ε)-secure PRP (not necessarily strong), then F is a $(t, q, \varepsilon + \frac{q^2}{2^{n+1}})$-secure PRF.

9 Secure symmetric encryption

Perfect secrecy is impossible where $m > \ell$, but computational security is possible with pseudorandom objects.

Electronic code block (ECB) Suppose F is a secure PRP $\{\mathtt{0},\mathtt{1}\}^\ell \times \{\mathtt{0},\mathtt{1}\}^n \to \{\mathtt{0},\mathtt{1}\}^n$ with F and F^{-1} efficiently computable.

Gen	$k \leftarrow \{\mathtt{0},\mathtt{1}\}^\ell$
Enc	$M' \leftarrow$ Pad message M with $\mathtt{1}$ and some $\mathtt{0}$s to a multiple of n. Break M' into n-bit blocks $m_0, m_1, \ldots$ Apply F_k to each of the $\{m\}$
Dec	Apply F'_k to each of the $\{m\}$

Repeated blocks give repeated ciphertext. Never use ECB.

Security model Adversary, seeing all cipthertexts and having oracle access to Enc_k, learns nothing about plaintexts (except message length, which is unavoidable).

$SE = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is (t, σ, ε)-IND-CPA secure ("indistinguishable under chosen-plaintext attack") iff $\forall$ (t, σ)-bounded $\mathcal{A}$,

$$\mathrm{Adv}_{SE}^{\mathrm{indcpa}}(\mathcal{A}) \equiv \Pr_{k \leftarrow \mathsf{Gen}}\left[\mathcal{A}^{L_k} \text{ accepts}\right] - \Pr_{k \leftarrow \mathsf{Gen}}\left[\mathcal{A}^{R_k} \text{ accepts}\right],$$

$$L_k(m, m') \equiv \mathsf{Enc}_k(m) \text{ if } |m| = |m'| \text{ else } \perp,$$
$$R_k(m, m') \equiv \mathsf{Enc}_k(m') \text{ if } |m| = |m'| \text{ else } \perp,$$

t is the running time, and σ total length of all message queries.

Equivalent definition: Enc_k is computationally indistinguishable from a zero-encrypting oracle $Z_k \equiv \mathsf{Enc}_k(\mathsf{0}^m)$.

Query repetition Enc in an IND-CPA-secure scheme should not always return the same ciphertext for multiple encryptions of the same message. This attack has advantage 1 against any deterministic and stateless scheme:

$c \leftarrow \mathcal{O}(\langle 0 \rangle, \langle 0 \rangle)$
$c' \leftarrow \mathcal{O}(\langle 0 \rangle, \langle 1 \rangle)$
Accept iff $c = c'$

10 Block cipher modes

Stateful counter mode (CTRS) Let F be a PRF with $m = n$.

Gen | $k \leftarrow \{\mathsf{0}, \mathsf{1}\}^\ell$, *counter* $\leftarrow 0$

Enc | echo *counter*
for each message block m:
  echo $F_k(\mathit{counter}) \oplus m_i$
  increment *counter*

CTRS is not used much, because preserving *counter* is difficult.

Theorem 4. If F is a (t, q, ε)-secure PRF, then CTRS(F) is $(t' \approx t, qn, 2\varepsilon)$-IND-CPA secure.

Proof. We will show using a hybrid argument that $\forall$ (t', σ)-bounded $\mathcal{A}'$ against CTRS(F) where $\sigma \leq n\,2^m$, there is a $(t \approx t', q = \sigma/n)$-bounded attacker $\mathcal{A}$ attacking F such that $\mathrm{Adv}_{\mathrm{CTRS}(F)}^{\mathrm{indcpa}}(\mathcal{A}') \leq 2\,\mathrm{Adv}_F^{\mathrm{prf}}(\mathcal{A})$.

Given $\mathcal{O}$ that is either F or U:

$$\mathcal{A} \equiv \mathcal{A}_L \equiv \left| \begin{array}{l} \mathit{counter} \leftarrow 0 \\ \mathcal{O}'(m, m') \equiv \left| \begin{array}{l} \text{If } |m| = |m'|, \text{ return } \perp \\ \text{Split } m \text{ into blocks } m_0, m_1, \ldots, m_{t-1} \\ y_i \leftarrow \mathcal{O}(\mathit{counter} + i) \ \forall\, i \in [0, t) \\ \text{Return } \mathit{counter} \| \mathrm{join}_i(m_i \oplus y_i) \\ \mathit{counter} \leftarrow \mathit{counter} + t \end{array} \right. \\ \text{Accept iff } \mathcal{A}'^{\mathcal{O}'} \text{ accepts} \end{array} \right.$$

Also define $\mathcal{A}_R$ similarly using m' instead of m.
$\mathcal{A}_L^{F_k}$ perfectly simulates L_k to $\mathcal{A}'$.
$\mathcal{A}_L^U$ does not simulate R_k, but it does simulate an oracle \$:

$$\$(m, m') \equiv \left| \begin{array}{l} \text{If } |m| = |m'|, \text{ return } \perp \\ \text{Return } \mathit{counter} \| [\text{random bits}] \\ \mathit{counter} \leftarrow \mathit{counter} + \text{number of blocks} \end{array} \right.$$

$$\begin{aligned} P_\ell &= \Pr_k[\mathcal{A}_L^{F_k}] = \Pr_k[\mathcal{A}'^{L_k}] \\ P_r &= \Pr_k[\mathcal{A}_R^{F_k}] = \Pr_k[\mathcal{A}'^{R_k}] \\ P_\$ &= \Pr_k[\mathcal{A}_L^U] = \Pr_k[\mathcal{A}_R^U] = \Pr_k[\mathcal{A}'^{\$}] \end{aligned}$$

$$\begin{aligned} \mathrm{Adv}^{\mathrm{indcpa}}_{\mathrm{CTRS}(F)}(\mathcal{A}') &= |P_\ell - P_r| \\ &\leq |(P_\ell - P_\$) + (P_r - P_\$)| \qquad \text{(triangle inequality)} \\ &\leq \varepsilon + \varepsilon = 2\varepsilon \end{aligned}$$

□

Counter modes

CTRS	One global counter	$\mathrm{adv}^{\mathrm{indcpa}} \leq 2\,\mathrm{adv}^{\mathrm{prf}}$
CTR\$	Random IV for each message	$\mathrm{adv}^{\mathrm{indcpa}} \leq 2\,\mathrm{adv}^{\mathrm{prf}} + q^2/2^n$
CTR\$\$	Random IV for each block	

Cipher block chaining (CBC) $C_0 = \mathrm{IV}$, $C_i = F_k(C_{i-1} \oplus m_i)$

Dec requires being able to calculate F^{-1}.

If F is a (t, q, ε)-secure PRF, then CBC$[F]$ is $(\approx t, \sigma = qn, 2\varepsilon + q^2/2^n)$-ind-cpa-secure. The proof requires showing that for U, all inputs to U are distinct (minus a birthday term).

11 Message authentication code (MAC)

Alice sends message m and $t \leftarrow \mathsf{Tag}_k(m)$. Eve intercepts (m, t) and delivers (m', t') to Bob. Bob runs $\mathsf{Ver}_k(m', t')$.

Ver_k returns $\begin{cases} m & \text{if } t' \text{ is a valid tag } (\mathsf{Ver}_k \text{ “accepts”}) \\ \perp & \text{otherwise } (\mathsf{Ver}_k \text{ “rejects”}) \end{cases}$

Eve has access to a Tag_k oracle and can make many attempts on Ver. Eve “wins” if Ver accepts on an m' not previously queried to Tag_k.

Conerns ignored by this model

Dropped messages

Replay attacks (“freshness” of messages)

Message sequence

Unforgeability under chosen message attack

$$\mathrm{Adv}^{\mathrm{ufmca}}_{\mathrm{MAC}}(\mathcal{A}) \equiv \Pr_{k \leftarrow \mathsf{Gen}}\left[\ \mathcal{A}^{\mathsf{Tag}_k, \mathsf{Ver}_k} \text{ “wins” }\right]$$

MAC is $(t, q_t, q_v, \varepsilon)$-uf-cma-secure iff advantage of an attacker bounded by time t, number of Tag queries q_t, and number of Ver queries q_v is less than ε.

Examples of reasonable constants

t	2^{80} or 2^{128}
q_v, q_t	2^{40} or 2^{56}
ε	2^{-40} or 2^{-56}

Brute-force MAC attacks

Key search: Get a few oracle tags, and guess k. $\mathrm{Adv} = t/2^{\ell}$.

Tag search: $\mathrm{Adv} = t/2^{s}$ where s is the tag length.

PRF-based MAC $\mathsf{Tag}_k \equiv F_k$

$\forall$ (t, q_t, q_v)-bounded $\mathcal{B}$, $\exists$ $(\approx t, q_t + q_v)$-bounded $\mathcal{A}$ such that

$$\mathrm{Adv}^{\mathrm{ufcma}}_{\mathrm{PRFMAC}[F]}(\mathcal{B}) \leq \mathrm{Adv}^{\mathrm{prf}}_{F}(\mathcal{A}) + q_v/2^n\ .$$

CBC-MAC For a fixed t, and $F : \{0,1\}^{nt} \to \{0,1\}^n$, CBC-MAC$[F]$ is secure, losing $(qt)^2/2^n$ advantage from that of F.

Cipher-based MAC (CMAC) Adds an extra step to the end of CBC-MAC to make it secure for arbitrary-length messages.

Precompute k_1, $k_2 \in \{\mathsf{0}, \mathsf{1}\}^n$ using $F_k(\mathsf{0}^m)$.

$$m'_t \leftarrow \begin{cases} m'_t \oplus k_1 & : |m'_t| = n \\ m' \| \mathsf{000} \ldots \oplus k_2 & : |m'_t| < n \end{cases}$$

Run $m_1 \| \ldots \| m_t$ through CBC-MAC.

12 Combining authenticity and privacy

Integrity of ciphertexts (INT-CTXT) $\mathsf{Dec}_k(c)$: returns decryption of c, or $\perp$ if c is invalid.

$SE = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is INT-CTXT secure iff $\forall$ bounded $\mathcal{A}$,

$$\mathrm{Adv}^{\text{int-ctxt}}_{SE}(\mathcal{A}) \equiv \Pr_{k \leftarrow \mathsf{Gen}} \left[\mathcal{A}^{\mathsf{Enc}_k, \mathsf{Dec}_k} \text{ wins} \right] < \varepsilon \ .$$

UF-CMA-security does not necessarily give INT-CTXT security. For example: If the output of Tag has a spurious bit that is ignored by Ver. So we require a stronger condition:

Strong unforgeability (SUF-CMA) Winning is redefined as: Ver_k accepts (m', t') that was not previously a query/answer pair to Tag_k.

Bad idea: Encrypt-and-tag $\mathsf{AEnc} \equiv \mathsf{EEnc}_{k_e}(m) \| \mathsf{Tag}_{k_m}(m)$. The tag could reveal information about m.

Bad idea: Tag-then-encrypt $\mathsf{AEnc} \equiv \mathsf{EEnc}_{k_e}(m \| \mathsf{Tag}_{k_m}(m))$. The ciphertext might be forgeable (for example, if EEnc appends a spurious bit).

Good idea: Encrypt-then-tag $\mathsf{AEnc} \equiv \mathsf{EEnc}_{k_e}(m) \| \mathsf{Tag}_{k_m}(\mathsf{EEnc}_{k_e}(m))$.

Indistinguishability under chosen ciphertext attack (IND-CCA) IND-CPA $\wedge$ INT-CTXT $\Rightarrow$ IND-CCA.

13 Summary of symmetric crypto games

		Oracles	Goal
Stronger ↓	IND-CPA IND-CCA	L/R L/R, $\hat{\mathsf{Dec}}$	Left or right?
Stronger ↓	INT-PTXT INT-CTXT	Enc, D	D returns new plaintext D returns non-$\perp$ on new ciphertext
Stronger ↓	UF-CMA SUF-CMA	Tag, Ver	$\mathsf{Ver}(m,t)$: m has never been tagged $\mathsf{Ver}(m,t)$: (m,t) has never been a Tag pair

14 Hashing

Hash function $h : D \to \{0,1\}^n$, $D > 2^n$

Collision $x, x' \in D : h(x) = h(x') \wedge x \neq x'$

Hash family $H : \{0,1\}^\ell \times D \to \{0,1\}^n$

Collision resistance (CR) H is (t, ε)-collision resistant if $\mathrm{Adv}_H^{\mathrm{cr}}(\mathcal{A}) \leq \varepsilon$ $\forall$ t-bounded $\mathcal{A}$.

$$\mathrm{Adv}_H^{\mathrm{cr}}(\mathcal{A}) = \Pr_{k \leftarrow \{0,1\}^\ell} [\ \mathcal{A}(k) \text{ outputs a collision in } H_k\ ]$$

Real-world hash functions

MD4	$n = 128$	Broken
MD5	$n = 128$	Broken
SHA-1	$n = 160$	Maybe broken
SHA-256	$n = 256$	Good
SHA-3		Good

Hash output lengths need to be longer than encryption key lengths because brute-force attacks can test $\approx q^2$ pairs in q hashes ("birthday attack").

Second preimage resistance / target collision resistance (TCR) Given x, attacker must find x' such that x, x' is a collision.

$$\mathrm{Adv}_H^{\mathrm{tcr}}(\mathcal{A}) = \Pr_{\substack{k \leftarrow \{0,1\}^\ell, \\ x \leftarrow D}} [\ \mathcal{A}(k, x) \text{ outputs a collision}\ ]$$

where D is some distribution over the message space.

Brute force attack does not have a birthday advantage in this game.

CR $\Rightarrow$ TCR.

One-wayness (OW)

$$\mathrm{Adv}_H^{\mathrm{ow}}(\mathcal{A}) = \Pr_{\substack{k \leftarrow \{0,1\}^\ell, \\ x \leftarrow D}} \left[\mathcal{A}(k, H_k(x)) \text{ outputs } x' : H_k(x') = H_k(x) \right]$$

TCR $\Rightarrow$ OW for "high-entropy" D (so that $H_k(x)$ reveals very little about x).

Merkle-Damgård (MD) transform $\mathrm{MD}[h]_k(M \in \{0,1\}^*)$

Uses a compression function $h_k : \{0,1\}^\ell \times \{0,1\}^{b+n} \to \{0,1\}^n$

Break M into $M_1, \ldots, M_t$ s.t. $\|M_i\| = b$

$$\begin{aligned}
y_1 &\leftarrow h_k(\ M_1 \,\|\, \langle 0 \rangle\ ) \\
y_2 &\leftarrow h_k(\ M_2 \,\|\, y_1\ ) \\
&\vdots \\
y_i &\leftarrow h_k(\ M_i \,\|\, y_{i-1}\ ) \\
&\vdots \\
y_t &\leftarrow h_k(\ M_t \,\|\, y_{t-1}\ ) \\
y &\leftarrow h_k(\ \langle t \rangle \,\|\, y_t\ )
\end{aligned}$$

If h is CR, then MD[h] is CR Let $\mathcal{B}$ attack $\mathrm{MD}[h]$ in CR game. We can build $\mathcal{A}$ attacking h in CR game. $\mathcal{A}(k \in \{0,1\}^\ell)$ will use a collision x, x' in $\mathrm{MD}[h_k]$ to find a collision in h_k.

If x, x' have different numbers of blocks: Then $\langle t \rangle \,\|\, M_t, \langle t' \rangle \,\|\, M_t'$ is a collision in h_k.

Otherwise: Walk backward through the MD process to find a step where $M_i \,\|\, y_{i-1} \neq M_i' \,\|\, y_{i-1}'$.

HMAC Secure MAC based on hash function.

$\mathrm{HMAC}_k(m) = H((k \oplus \mathrm{opad}) \,\|\, H((k \oplus \mathrm{ipad}) \,\|\, m))$ where k is the MAC key padded to the length of the compression function, ipad and opad are fixed strings of the same length.

Only TCR security of H is required for the security of HMAC.

15 Groups

Group axioms

G is closed under $\cdot$

$\mathbb{1} \cdot a = a \cdot \mathbb{1} = a$

Every element of G has a unique multiplicative inverse

$(a \cdot b) \cdot c = a \cdot (b \cdot c)$

The group operator $\cdot$ is not necessarily commutative.

Examples of groups

Integers under addition

Invertible matrices with real entries under multiplication

$\mathbb{Z}_N = \{0, 1, \ldots, N-1\}$ under modular addition

$\mathbb{Z}_N^* = \{a \in \mathbb{Z}_N : \gcd(a, N) = 1\}$ under multiplication $\pmod N$

$\mathbb{Z}_N^*$ is closed under $\cdot$ Proof sketch:
$\gcd(a \cdot b, N) = 1 \implies \gcd(ab \bmod N, N) = 1$.

$\mathbb{Z}_N^*$ has unique inverses Proof sketch:
Use the fact $\gcd(a, N) = 1 \iff \exists\, x, y \in \mathbb{Z} : ax + Ny = 1$.
$ax = 1 - Ny = 1 \mod N \implies x = a^{-1} \mod N$.

Generator $g \in G : \{g^0, g^1, g^2, \ldots\} = G$
If G is finite, then $a^{|G|} = 1 \; \forall\, a \in G$.
g is a generator iff $g^i \neq \mathbb{1} \; \forall\, 0 < i < |G| : i \mid G$
For a prime p, $\mathbb{Z}_p^*$ is finite and cyclic (has a generator). Its order is $p-1$.

Order of $a \in G$ is $(\min i > 0 : a^i = 1)$.
$\text{order}(a)$ divides $|G|$.

How many generators? If $g \in G$ is a generator and i is coprime with $|G|$, then g^i is a generator. Therefore we have $\left|\mathbb{Z}_{p-1}^*\right|$ generators.

16 Diffie-Hellman protocol (1976)

For establishing a secret key without meeting.

A chooses $a \in G$ and transmits g^a. B chooses $b \in G$ and transmits g^b. The secret is g^{ab}.

Requires a multiplicative group $(G, \cdot)$ and a generator $g \in G$, and the discrete log problem must be intractable.

Calculations required for implementation Multiply huge numbers ($\approx 2^{4096}$), exponentiate huge numbers, sample huge primes, find a generator of $\mathbb{Z}_p^*$.

Multiplication Simple grade-school multiplication algorithm followed by Euclid's division algorithm. Both are quadratic in bit length.

Exponentiation Repeated squaring. $O(\lg^3 a)$ multiplications.

Primality testing Miller-Rabin test runs in $O(\lg^3 p)$, randomized with false positives. Alternately, AKS is a deterministic test running in $O(\lg^6 p)$. Approximately $1/k$ of k-bit numbers are prime.

Generator testing g is a generator if $g^{(p-1)/i} = 1\ \forall$ prime divisors i of $p-1$. Since factoring is hard, we must generate p such that we know the factors of $p-1$ ...

Sophie Germain (safe) primes Pick a prime q and let $p = 2q + 1$. Repeat until p is prime. It appears that $1/k^2$ of k-bit numbers are safe primes, although this is unproven.

Uniformity of shared secret We want $ab \bmod (p-1)$, and therefore also g^{ab}, to be uniformly distributed. If $G = \mathbb{Z}_p^*$, this is not the case; ab is even with probability $3/4$, and the $\bmod\,(p-1)$ operation does not affect this parity (because $p-1$ is even). But if G has prime order q, then $ab \bmod q$ is very nearly uniform ($ab = 0$ with probability $2/q$, anything else with probabilility $1/(q-1)$).

Quadratic residue The subgroup of quadratic residues in $\mathbb{Z}_p^*$ is $\mathbb{QR}_p^* = \{(g^2)^0, (g^2)^1, (g^2)^2, \ldots, (g^2)^{(p-3)/2}\}$. For $p = 2q+1$, $\left|\mathbb{QR}_p^*\right| = q$, so if p is a safe prime, $\left|\mathbb{QR}_p^*\right|$ has prime order. g^2 is a generator for $\mathbb{QR}_p^*$.

Jacobi symbol For $y \in \mathbb{Z}_p^*$, Jacobi symbol is 1 if $y \in \mathbb{QR}_p^*$, and -1 otherwise. $y \in \mathbb{QR}_p^* \iff y^{(p-1)/2} = 1$. Proof: $y^{(p-1)/2} = g^{i(p-1)/2}$ for some $i = 2i' + b$, $b \in \{0,1\}$. Use the fact that $g^{(p-1)/2} = -1$. $y = g^{b(p-1)/2} = \{1 \text{ if } b = 0,\ -1 \text{ if } b = 1\}$.

17 Asymmetric encryption

Scheme AE = (Gen, Enc, Dec)

$(\mathrm{pk}, \mathrm{sk}) \leftarrow \mathsf{Gen}$

$c \leftarrow \mathsf{Enc}_{\mathrm{pk}}(m) = \mathsf{Enc}(\mathrm{pk}, m)$ where $m \in M_{\mathrm{pk}}$, and M_{pk} is some group

m or $\perp \leftarrow \mathsf{Dec}_{\mathrm{sk}}(c)$

Gen and Enc are randomized. Dec is deterministic.

IND-CPA-security

$$\mathrm{Adv}_{\mathrm{AE}}^{\mathrm{indcpa}}(\mathcal{A}) \equiv \Pr_{(\mathrm{pk},\mathrm{sk}) \leftarrow \mathsf{Gen}}\left[\mathcal{A}^{L_{\mathrm{pk}}}(\mathrm{pk})\right] - \Pr_{(\mathrm{pk},\mathrm{sk}) \leftarrow \mathsf{Gen}}\left[\mathcal{A}^{R_{\mathrm{pk}}}(\mathrm{pk})\right]$$

where $L_{\mathrm{pk}}(m_L, m_R) = \mathsf{Enc}_{\mathrm{pk}}(m_L)$ and $R_{\mathrm{pk}}(m_L, m_R) = \mathsf{Enc}_{\mathrm{pk}}(m_R)$.

IND-CCA-security Like IND-CPA, but the attacker gets another oracle:

$$D_{\mathrm{sk}}(c) = \begin{cases} \mathsf{Dec}_{\mathrm{sk}}(c) & \text{if } c \text{ was not returned by L/R oracle} \\ \perp & \text{otherwise} \end{cases}.$$

One encryption query $\forall$ ind-cpa attacker $\mathcal{B}$ against AE, $\exists$ an ind-cpa attacker $\mathcal{A}$ against AE making $\leq q$ Enc queries such that $\mathrm{Adv}_{\mathrm{AE}}^{\mathrm{indcpa}}(\mathcal{B}) \leq q\,\mathrm{Adv}_{\mathrm{AE}}^{\mathrm{indcpa}}(\mathcal{A})$. (Also, same for ind-cca with 1 encryption query and unlimited D_{sk} queries.) The analogous claim is not true in a symmetric key setting.

ElGamal Asymmetric encryption scheme similar to Diffie-Hellman.

Gen: Choose group G of order n, generator g. Choose $x \leftarrow \mathbb{Z}_n$. Output $(\mathrm{sk} = x, \mathrm{pk} = (G, G, X = g^x))$.

$\mathsf{Enc}_{\mathrm{pk}}(m \in G)$: Choose $y \leftarrow \mathbb{Z}_n$. Output $c = (Y = g^y, \overbrace{m \cdot X^y}^{g^{xy}})$.

$$\mathsf{Dec}_{\text{sk}}(\overbrace{Y}^{g^y}, \overbrace{Z}^{m\cdot g^{xy}})\text{: Output } \overbrace{Z}^{m\cdot g^{xy}} \cdot (\overbrace{Y^x}^{g^{xy}})^{-1} = m.$$

Not ind-cpa-secure for $G = \mathbb{Z}_p^*$. Attack: $(Y, Z) \leftarrow O(1, g)$. Accept if $Z^{(p-1)/2} = 1$, reject if -1. $\Pr[\mathcal{A}^L] = \Pr[1 \cdot g^{xy} \text{ is square}] = 3/4$. $\Pr[\mathcal{A}^R] = \Pr[g^{xy+1} \text{ is square}] = 1/4$.

Is ind-cpa-secure under assumption that DDH (decision Diffie-Hellman) is hard.

Is not cca-secure. But the Cramer-Shoup scheme is, using two applications of ElGamal, under the DDH assumption.