One-time Pad

- The *XOR* operation: $a \oplus b = a + b \bmod 2$.

  $a \oplus 0 = a$
  $a \oplus a = 0$
  $a \oplus b = b \oplus a$ (Commutativity)
  $a \oplus (b \oplus c) = (a \oplus b) \oplus c$ (Associativity)

  The One-time Pad scheme (Vernam 1917, Shannon 1949):
  $n = |k| = |x|$, $Enc : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$
  $Enc_k(x) = x \oplus k$
  $Dec_k(y) = y \oplus k$

  Validity:
  $Dec_k(Enc_k(x)) = (x \oplus k) \oplus k = x \oplus (k \oplus k) = x \oplus 0^n = x$

一次一密的密钥k不能重复使用
因为若 $C_1 = k \oplus x_1$, $C_2 = k \oplus x_2$, 则 $C_1 \oplus C_2 = (k \oplus x_1) \oplus (k \oplus x_2) = x_1 \oplus x_2$
泄露了明文信息

$\mathcal{Q}$: What about if using the same key twice?
Say $c_1 = k \oplus m_1$ and $c_2 = k \oplus m_2$.
$c_1 \oplus c_2 = (k \oplus m_1) \oplus (k \oplus m_2) = m_1 \oplus m_2$
$m_1 \oplus m_2$ leaks information about m_1, m_2

- **Theorem 1.9** One-time pad is *perfectly secure*.

  Proof. Prove that for every $x \in \{0,1\}^n$, the distribution $Y_x = Enc_{U_n}(x)$ is uniformly distributed.

  Let $y \in \{0,1\}^n$, we need to show that
  $Pr_{k \leftarrow_R \{0,1\}^n}[x \oplus k = y] = 2^{-n}$
  Since there is a unique single value of $k = x \oplus y$, the probability that the equation is true is 2^{-n}.

  Proof. Fix arbitrary distributuion over $\mathcal{M} = \{0,1\}^n$, and arbitrary $m, c \in \{0,1\}^n$.

  $\Pr[M = m \mid C = c] = ?$
  $= \Pr[C = c \mid M = m] \cdot \Pr[M = m] / \Pr[C = c]$
  $= \Pr[K = m \oplus c] \cdot \Pr[M = m] / 2^{-n}$
  $= 2^{-n} \cdot \Pr[M = m] / 2^{-n}$
  $= \Pr[M = m]$

  $\Pr[C = c]$
  $= \sum_{m'} \Pr[C = c \mid M = m'] \cdot \Pr[M = m']$
  $= \sum_{m'} \Pr[K = m' \oplus c] \cdot \Pr[M = m']$
  $= \sum_{m'} 2^{-n} \cdot \Pr[M = m']$
  $= 2^{-n}$

一次一密是完全安全的

当x与y固定时，密钥 $k = x \oplus y$ 是固定的
由于 $x, y \in \{0,1\}^n$，故k取k'的概率为 2^{-n}

Limitation of perfect secrecy

- $\mathcal{Q}$: Is this the end of cryptography?

  We need more:
  ◇ Use the same key for many plaintexts
  ◇ Use n-bit key for $2n$-bit plaintexts.

  Theorem 1.10 (Limitations of perfect secrecy) There is no *perfectly secure* encryption schemes (Gen, Enc, Dec) with n-bit plaintexts and $(n-1)$-bit keys.

  Proof. See blackboard.

  Suppose that (Gen, Enc, Dec) is such an encryption scheme. Denote by Y_0 the distribution $Enc_{U_{n-1}}(0^n)$ and by S_0 its support. Since there are only 2^{n-1} possible keys, we have $|S_0| \le 2^{n-1}$.
  Now for every key k the function $Enc_k(\cdot)$ is one-to-one and hence its image is of size $\ge 2^n$. This means that for every k, there exists x such that $Enc_k(x) \notin S_0$. Fix such a k and x, then the distribution $Enc_{U_{n-1}}(x)$ does not have the same support as Y_0 and hence it is not identical to Y_0.

  - The key is as long as the message
  - Only secure if each key is used to encrypt a single message
  - Trivially broken by a known-plaintext attack

  Whenever faced with an impossibility result, it is a good idea to examine whether we can relax these assumptions to still get what we want (or at least something close to that).

完美安全要求密文长度至少为明文长度.

只需证明 $Enc_{U_{n-1}}(x) \neq Enc_{U_{n-1}}(x')$
令 $Y_0 = Enc_{U_{n-1}}(0^n)$, S_0 为 Y_0 的支撑集(即 $Y_0 > 0$ 的取值)
因为密钥只有 2^{n-1} 种取值，故 $|S_0| \le 2^{n-1}$.
$\forall k$, 要求 $Enc_k(\cdot)$ 是单射. 因此映射空间大小 $\ge 2^n$
所以 $\forall k, \exists x \in \{0,1\}^n$ s.t. $Enc_k(x) \notin S_0$.
此时 $Enc_{U_{n-1}}(x)$ 没有与 Y_0 相同的支撑集，即 $Enc_{U_{n-1}}(x) \neq Y_0$

Key generation

- When describing algorithms, we assume access to uniformly distributed bits/bytes. Where do these actually come from?

- Precise details depend on the system
  - Linux or unix: /dev/random or /dev/urandom
  - Do not use *rand()* or *java.util.Random*
  - Use crypto libraries instead

Random-number generation

- Two steps:
  1. Continually collect a "pool" of high-entropy ("unpredictable") data
     - Must ultimately come from some physical process (keystroke/mouse movements, delays between network events, hard-disk access times, hardware random-number generation (Intel), etc.)
  2. (Smoothing) When random bits are requested, process this data to generate a sequence of uniform, independent bits/bytes
     - Need to eliminate both *bias* and *dependencies*

较稳健的随机数一般与物理过程有关.

- Step 2: Smoothing

  von Neumann technique for eliminating *bias*:
  - Collect two bits per output bit
    - $01 \to 0$, $10 \to 1$, $00, 11 \to$ skip
  - Note that this assumes *independence* (as well as constant bias)

- Read desired number of bytes from "/dev/urandom"

ϵ-Statistical Security

- Eve has a tiny advantage in its posteriori guessing probability compared to the priori probability.

  We may say that an encryption scheme is *ϵ-statistically indistinguishable* if the probability that Eve guesses which of the two messages was encrypted is at most $1/2 + \epsilon$.

  $\mathcal{Q}$:
  1) Is the relaxed definition still strong enough in practice?
  2) Does the relaxation buy something we could not get with the original definition (perfect security)?

  $\mathcal{A}$:
  1) Yes, if ϵ is small (say 10^{-6} or even 10^{-100})
  2) No, we cannot have key shorter than the message.

放缩了ϵ

此时完全安全仍要求密钥不短于明文.

- **Definition 2.1** Let X and Y be two distributions over $\{0,1\}^n$. The *statistical distance* of X and Y, denoted by $\Delta(X, Y)$ is defined to be
  $\max_{T \subseteq \{0,1\}^n} |\Pr[X \in T] - \Pr[Y \in T]|$.
  If $\Delta(X, Y) \leq \epsilon$, we say that $X \equiv_\epsilon Y$.

  Definition 2.2 *ϵ-Statistical Security*. An encryption scheme (Gen, Enc, Dec) is *ϵ-statistically secure* if for every pair of plaintexts m, m', we have $Enc_{U_n}(m) \equiv_\epsilon Enc_{U_n}(m')$.

最大概率距离 $\leq \epsilon$.

- **Lemma 2.3**

$$\Delta(X, Y) = \frac{1}{2} \sum_{w \in Supp(X) \cup Supp(Y)} |Pr[X = w] - Pr[Y = w]|$$

  Proof. See blackboard.

  Observations:
  $0 \leq \Delta(X, Y) \leq 1$
  $\Delta(X, Y) = 0$ if $X = Y$
  $0 \leq \Delta(X, Y) \leq \Delta(X, Z) + \Delta(Z, Y)$

  Δ is a *metric*.

对任意集合T，定义 $\Delta_T(X,Y) = |\Pr[X\in T] - \Pr[Y\in T]|$

则 $\Delta(X,Y) = \max_{T\subseteq\{0,1\}^n} \Delta_T(X,Y)$

① 因为 $\Pr[X\in T^c] = 1 - \Pr[X\in T]$，有 $\Delta_{T^c}(X,Y) = \Delta_T(X,Y)$

令 $T = \{w \mid \Pr[X=w] > \Pr[Y=w]\}$，则

$$\frac{1}{2}\sum_{w\in Supp(X)\cup Supp(Y)} |\Pr[X=w]-\Pr[Y=w]| = \frac{1}{2}\sum_{w\in T}|\cdots| + \frac{1}{2}\sum_{w\in T^c}|\cdots|$$

$$= \frac{1}{2}\sum_{w\in T}(\Pr[X=w]-\Pr[Y=w]) + \frac{1}{2}\sum_{w\in T^c}(\Pr[X=w]-\Pr[Y=w]) = \frac{1}{2}\Delta_T(X,Y) + \frac{1}{2}\Delta_{T^c}(X,Y)$$

$$= \Delta_T(X,Y) \leq \Delta(X,Y)$$

② 设集合S使得 $\Delta_S(X,Y)$ 取得最大值，即 $\Delta(X,Y) = \Delta_S(X,Y)$

不失一般性，可令 $\Pr[X\in S] \geq \Pr[Y\in S]$（若不满足，则取 $S = S^c$）

$$2\Delta(X,Y) = \Delta_S(X,Y) + \Delta_{S^c}(X,Y) = \Pr[X\in S] - \Pr[Y\in S] + \Pr[Y\in S^c] - \Pr[X\in S^c]$$

$$= \sum_{w\in S}(\Pr[X=w]-\Pr[Y=w]) + \sum_{w\in S^c}(\Pr[Y=w]-\Pr[X=w])$$

$$\leq \sum_{w\in S}|\Pr[X=w]-\Pr[Y=w]| + \sum_{w\in S^c}|\Pr[Y=w]-\Pr[X=w]| = \sum_w |\Pr[X=w]-\Pr[Y=w]|$$

（$a+b \leq |a|+|b|$）

所以 $\Delta(X,Y) \leq \frac{1}{2}\sum_w |\Pr[X=w]-\Pr[Y=w]|$

综上：$\Delta(X,Y) = \frac{1}{2}\sum_{w\in\cdots} |\Pr[X=w]-\Pr[Y=w]|$

metric 三性质：① $\delta(x,x)=0$, $\forall x$

② $\delta(x,y)=\delta(y,x)$, $\forall x,y$

③ $\delta(x,z) \leq \delta(x,y)+\delta(y,z)$, $\forall x,y,z$

Lemma 2.4 Eve has at most $1/2+\epsilon$ success probability if and only if for every pair of m_1, m_2,
$\Delta(Enc_{U_n}(m_1), Enc_{U_n}(m_2)) \le 2\epsilon$.

Proof.
Suppose that Eve has $1/2+\epsilon$ success probability with m_1, m_2. Let $p_{i,j} = \Pr[Eve(Enc_{U_n}(m_i)) = j]$. Then we have

$$p_{1,1} + p_{1,2} = 1$$
$$p_{2,1} + p_{2,2} = 1$$
$$(1/2)p_{1,1} + (1/2)p_{2,2} \le 1/2 + \epsilon.$$

The last two together imply that

$$p_{1,1} - p_{2,1} \le 2\epsilon,$$

which means that if we let T be the set $\{c : Eve(c) = 1\}$, then T demonstrates that $\Delta(Enc_{U_n}(m_1), Enc_{U_n}(m_2)) \le 2\epsilon$.
Similarly, if we have such a set T, we can define an attacker from it that succeeds with probability $1/2+\epsilon$.

因为由ε-statistically indistinguishable定义
$p_{1,1} \le \frac{1}{2}+\epsilon$
$p_{2,2} \le \frac{1}{2}+\epsilon$
$\Rightarrow \frac{1}{2}p_{1,1} + \frac{1}{2}p_{2,2} \le \frac{1}{2}+\epsilon$.

Limitation of ε-Statistical Security

Theorem 2.5 Let (Gen, Enc, Dec) be a valid encryption with $Enc : \{0,1\}^n \times \{0,1\}^{n+1} \to \{0,1\}^*$. Then there exist plaintexts m_1, m_2 with $\Delta(Enc_{U_n}(m_1), Enc_{U_n}(m_2)) > 1/2$.

Proof. See blackboard.

Fact. For a random variable Y, if $E[Y] \le \mu$ the $\Pr[Y \le \mu] > 0$.
Let $m_1 = 0^{n+1}$, and let $S = Supp(Enc_{U_n}(m_1))$, then $|S| \le 2^n$.
We choose a random message $m \leftarrow_R \{0,1\}^{n+1}$ and define the following 2^n random variables for every k:

$$T_k(m) = \begin{cases} 1, & \text{if } Enc_k(m) \in S \\ 0, & \text{otherwise} \end{cases}$$

Since for every k, $Enc_k(\cdot)$ is one-to-one, we have $\Pr[T_k = 1] \le 1/2$. Define $T = \sum_{k\in\{0,1\}^n} T_k$, then
$E[T] = E[\sum_k T_k] = \sum_k E[T_k] \le 2^n/2$.
This means the probability $\Pr[T \le 2^n/2] > 0$. In other words, there exists an m s.t. $\sum_k T_k(m) \le 2^n/2$. For such m, at most half of the keys k satisfy $Enc_k(m) \in S$, i.e.,
$\Pr[Enc_{U_n}(m) \in S] \le 1/2$.
Since $\Pr[Enc_{U_n}(0^{n+1}) \in S] = 1$, we have
$\Delta(Enc_{U_n}(0^{n+1}), Enc_{U_n}(m)) > 1/2$.

等价于：若$|k|<|m|$，则有ε-statistical security的加密机制不存在

事实：对r.v. Y，若$E[Y]\le\mu$，则$\Pr[Y\le\mu]>0$

令$m_1 = 0^{n+1}$，$S = Supp(Enc_{U_n}(m_1))$，则$|S| \le 2^n$.

令$m \in_R \{0,1\}^{n+1}$，且对任意k定义2^n个r.v.的取值.

$$T_k(m) = \begin{cases} 1, & \text{if } Enc_k(m) \in S \\ 0, & \text{otherwise} \end{cases}$$

对任意k，$Enc_k(\cdot)$是单射的，则$\Pr[T_k(m)=1] \le \frac{1}{2}$ ①

定义$T = \sum_{k\in\{0,1\}^n} T_k$，则由期望的线性性，有

$$E[T] = E[\sum_{k\in\{0,1\}^n} T_k] = \sum_{k\in\{0,1\}^n} E[T_k] = \sum_{k\in\{0,1\}^n} \left[1\cdot\Pr[T_k(m)=1] + 0\cdot\Pr[T_k(m)=0]\right]$$
$$= \sum_{k\in\{0,1\}^n} \Pr[T_k(m)=1] \le 2^n \cdot \frac{1}{2}$$

由事实可知：$\Pr[T \le \frac{2^n}{2}] > 0$.

即存在m有$T = \sum_{k\in\{0,1\}^n} T_k(m) \le \frac{2^n}{2}$

对这样的m，有$\Pr[Enc_{U_n}(m) \in S] \le \frac{1}{2}$ ②

而$\Pr[Enc_{U_n}(0^{n+1}) \in S] = 1$，

故$\Delta(Enc_{U_n}(0^{n+1}), Enc_{U_n}(m)) = \max_{T\subseteq\{0,1\}^n} |P[X\in T] - P[Y\in T]|$
$> |1 - \frac{1}{2}| = \frac{1}{2}$

证①：因为$m \in_R \{0,1\}^{n+1}$，故$|Enc_k(m)| = 2^{n+1}$
而$|S| \le 2^n$
所以取$Enc_k(m) \in S$最多也只有$\frac{1}{2}$
即$\Pr[T_k(m)=1] \le \frac{1}{2}$

证②：有$T_k(m) = \begin{cases} 1, & \text{if } Enc_k(m) \in S \\ 0, & \text{otherwise} \end{cases}$
且$\sum_{k\in\{0,1\}^n} T_k(m) \le \frac{2^n}{2}$，$|k| = 2^n$
若$T_k(m)$全取1，则$\sum T_k(m) = 2^n$
故$\Pr[T_k(m)=1] \le \frac{1}{2}$，即最多只有一半的$T_k(m)$取1.
所以，$\Pr[Enc_k(m) \in S] \le \frac{1}{2}$

Computational Security 计算安全

- Statistical security does not allow us to break the impossibility result.
  - In real life, people are using encryption with keys shorter than the message size to encrypt all kinds of sensitive information.
  - If the algorithm you use to break the encryption scheme runs in time 2^n , it seems OK since the message may be expired by then ...
- **Idea**: Would be OK if a scheme leaked information with *tiny probability* to eavesdroppers with *bounded computational resources*
  - Allowing security to "fail" with tiny probability
  - Restricting attention to "efficient" attackers

密钥需要短于明文

攻击者能力有限

- Two problems:

1) While the particular algorithm runs in exponential time, we cannot guarantee that there is no other algorithm is efficient.

e.g., the substitution cipher has a huge key space, but can be broken efficiently.

2) We need a precise mathematical definition (like *prefect secrecy* definition).

"Breaking E is very hard"?
"Problem P cannot be solved in reasonable time"?

Q: How do we model the resources of Eve (the adversary)?

Perfect indistinguishability 完美不可区分性

- **Definition 1.6** *Perfect secrecy*. An encryption scheme (Gen, Enc, Dec) with message space $\mathcal{M}$ and ciphertext space $\mathcal{C}$ is *perfectly secure* if and only if for every two distinct plaintexts $\{m_0, m_1\} \in \mathcal{M}$, and for every strategy used by Eve, if we choose at random $b \in \{0,1\}$ and a random key $k \in \{0,1\}^n$, then the probability that Eve guesses m_b after seeing the ciphertext $c = Enc_k(m_b)$ is at most $1/2$.

- Let $\Pi = (Gen, Enc, Dec)$ be an enryption scheme with message space $\mathcal{M}$, and A an adversary

  Define a randomized experiment $PrivK_{A,\Pi}$:
  1. A outputs $m_0, m_1 \in \mathcal{M}$
  2. $k \leftarrow Gen$, $b \leftarrow \{0,1\}$, $c \leftarrow Enc_k(m_b)$
  3. $b' \leftarrow A(c)$

  Adversary A *succeeds* if $b = b'$, and we say the experiment evaluates to 1 in this case.

  Π is *perfectly indistinguishable* if for all attackers (algorithms) A, it holds that $\Pr[PrivK_{A,\Pi} = 1] \leq 1/2$

  Claim: Π is *perfectly indistinguishable* $\Leftrightarrow$ Π is *perfectly secure*

完美不可区分性与完美安全是等价的.

- **Idea**: relax *perfect indistinguishability*

  Two approaches
  - *Concrete* security
  - *Asympotic* security

具体化 (确定ε的值)
渐近化 (确定$\varepsilon = 2^{-n}$)

Computational indistinguishability 计算不可区分性

- (t, ϵ)-*indistinguishability* (concrete)
  - Security may fail with probability $\leq \epsilon$
  - Restrict attention to attackers running in time $\leq t$
- Π is (t, ϵ)-indistinguishable if for all attackers A running in time at most t, it holds that
$$\Pr[PrivK_{A,\Pi} = 1] \leq 1/2 + \epsilon$$

Does not lead to a clean theory ...
- Sensitive to exact computational model
- Π can be (t, ϵ)-secure for many choices of t, ϵ

- Introduce *security parameter* n (asymptotic)
  - For now, can view n as the key length
  - Fixed by honest parties at initialization
  - Known by adversary

Measure running time of all parties, and the success probability of the adversary, as functions of n

引入安全参数 n

Computational indistinguishability:
- Security may fail with probability *negligible* in n
- Restrict attention to attackers running in time (at most) polynomial in n

- A function $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ is (at most) *polynomial* if there exists c s.t. $f(n) < n^c$ for large enough n.

A function $f : \mathbb{Z}^+ \to [0, 1]$ is *negligible* if every polynomial p it holds that $f(n) < 1/p(n)$ for large enough n.
  - Typical example: $f(n) = poly(n) \cdot 2^{-cn}$
- "Efficient" = "(probabilistic) polynomial-time (**PPT**)" borrowed from *complexity theory*
- Convenient closure properties
  - $poly * poly = poly$
    - Poly-many calls to PPT subroutine (with poly-size input) is still PPT
  - $poly * negl = negl$
    - Poly-many calls to subroutine that fails with negligible probability fails with negligible probability overall

可忽略函数：对多项式 p，有 $f(n) < \frac{1}{p(n)}$ 对足够大的 n.

- A *private-key encryption scheme* is defined by three PPT algorithms (Gen, Enc, Dec):
  - Gen: takes as input 1^n; outputs k
  - Enc: takes as input a key k and message $m \in \{0,1\}^*$; outputs ciphertext c: $c \leftarrow Enc_k(m)$
  - Dec: takes key k and ciphertext c as input; outputs a message m or "error" ($\perp$)

私钥加密模式中的算法都是PPT算法.

Computational indistinguishability (asymptotic)

- Fix Π, A
  Define a randomized experiment $PrivK_{A,\Pi}(n)$:
  1. $A(1^n)$ outputs $m_0, m_1 \in \{0,1\}^*$ of equal length
  2. $k \leftarrow Gen(1^n)$, $b \leftarrow \{0,1\}$, $c \leftarrow Enc_k(m_b)$
  3. $b' \leftarrow A(c)$

  Adversary A *succeeds* if $b = b'$, and we say the experiment evaluates to 1 in this case.

  Definition 3.1 Π is *computationally indistinguishable* (aka *EAV-secure*) if for all PPT attackers (algorithms) A, there is a *negligible* function ϵ such that
  $$\Pr[PrivK_{A,\Pi}(n) = 1] \leq 1/2 + \epsilon(n)$$

EAV: eavesdropper
窃听者只看到密文(也即已知密文攻击)

- Consider a scheme where the best attack is *brute-force search* over the key space, and $Gen(1^n)$ generates a uniform n-bit key
  - So if A runs in time $t(n)$, then
    $$\Pr[PrivK_{A,\Pi}(n) = 1] < 1/2 + O(t(n)/2^n)$$
  - The scheme is EAV-secure: for any polynomial t, the function $t(n)/2^n$ is negligible.

①限定了攻击算法是negligible的
②Scheme用PPT算法.

Encryption and plaintext length

- In practice, we want encryption schemes that can encrypt arbitrary-length messages.
- In general, encryption does not hide the plaintext length
  - The definition takes this into account by requiring m_0, m_1 to have the same length.
- But leaking plaintext length can often lead to problems in the real world!
  - Databases searches
  - Encrypting compressed data

现实中希望encryption scheme能加密任意长信息