完善保密加密机制

Perfectly-secret Encryption

安全目标

无条件安全（Unconditional Security）

即使攻击者具有无限的计算资源，也无法攻破密码体制，我们称这种密码体制是无条件安全的。

无条件安全又称为完善保密：Perfect Secrecy

加密算法定义

加密算法包括三个子算法：

1. 密钥生成子算法(Gen)

   算法输入：安全参数n；算法输出：满足特定分布的密钥k

2. 加密子算法(Enc)

   算法输入：密钥k和明文m；算法输出：密文c，c=Enc(k, m)

3. 解密子算法(Dec)

   算法输入：密钥k和密文c；算法输出：明文m，m=Dec(k, c)

无条件安全定义

一个加密算法（Gen，Enc，Dec）在明文空间M上是无条件安全的，如果对于M的每种概率分布，其中的每个消息m和每个可能出现的密文c，均满足如下条件：

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

解释例子

	a (1/4)	b (3/4)
K_1 (1/2)	1	2
K_2 (1/4)	2	3
K_3 (1/4)	3	4

$$\Pr[a] = \frac{1}{4}$$

$$\Pr[b] = \frac{3}{4}$$

$$\Pr[a \mid 1] = 1$$

$$\Pr\left[a \middle| 2\right] = \frac{1}{7}$$

$$\Pr[\mathrm{a} \mid 3] = \frac{1}{4}$$

$$\Pr[\mathrm{a} \mid 4] = 0$$

$$\Pr[b \mid 1] = 0$$

$$\Pr\left[b \middle| 2\right] = \frac{6}{7}$$

$$\Pr[\mathrm{b} \mid 3] = \frac{3}{4}$$

$$\Pr[\mathrm{b} \mid 4] = 1$$

$$\Pr\left[1\right] = \frac{1}{8}$$

$$\Pr\left[2\right] = \frac{3}{8} + \frac{1}{16} = \frac{7}{16}$$

$$\Pr\left[3\right] = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}$$

$$\Pr\left[4\right] = \frac{3}{16}$$

解释例子

- 移位密码是无条件安全的吗？ （若字母出现频率相同）
  - 移位密码

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

  - 单字母替换

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
X	E	U	A	D	N	B	K	V	M	R	O	C	Q	F	S	Y	H	W	G	L	Z	I	J	P	T

解释例子

1） 凯撒密码：

$$M = C = Z_{26}^{1}, K = Z_{26}^{1}$$

$$\Pr[\mathrm{M} = 'A'] = \tfrac{1}{26}$$

$$\Pr[\mathrm{M} = 'A' \mid C = 'D'] = \Pr[\mathrm{K} = 3] = \tfrac{1}{26}$$

$$\Pr[\mathrm{M} = 'AA' \mid C = 'CD'] = 0$$

2） 单字母替换：

$$M = C = Z_{26}^{5}$$

$$\Pr[\mathrm{M} = 'ABCDE'] = \tfrac{1}{26^5}$$

$$\Pr[\mathrm{M} = 'ABCDE' \mid C = 'EECBA'] = 0$$

解释例子

3）维吉尼亚密码 $M = C = K = Z_{26}^{5}$

$$\Pr[\mathrm{M} = 'ABCDE'] = \frac{1}{26^5}$$

$$\Pr[\mathrm{M} = 'ABCDE' \mid C = 'EDCBA'] = \frac{1}{26^5}$$

$$\Pr[\mathrm{M} = 'ABCDE' \mid C = 'EEEEE'] = \frac{1}{26^5}$$

先确定密钥长度，才能找到反例。

如密钥长为5，则：

Pr[M = ‘ABCDEA’ | C=‘EEEEEF’] = 0

上述概率均为唯密文攻击下的概率，如果针对已知明文攻击等，无法达到无条件安全的反例更容易找到。

练习

一个加密算法（Gen，Enc，Dec）在明文空间M上是无条件安全的，如果对于M的每种概率分布，其中的每个消息m和每个可能出现的密文c，均满足如下条件：

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

$$Pr[C = c \mid M = m] = Pr[C = c]$$

成立?

练习

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

$$\frac{\Pr[M = m \wedge C = c]}{\Pr[C = c]} = \Pr[M = m]$$

$$\frac{\Pr[C = c \wedge M = m]}{\Pr[M = m]} = \Pr[C = c]$$

$$\Pr[C = c \mid M = m] = \Pr[C = c]$$

定理

一个加密算法（Gen，Enc，Dec）在明文空间M上是无条件安全的，**当且仅当**对于M的每种概率分布，其中的任意两个消息m0和m1，和每个可能出现的密文c，如下条件均满足：

$$\Pr[C = c \mid M = m0] = \Pr[C = c \mid M = m1]$$

充分性

$$\frac{\Pr[M=m0]}{\Pr[M=m0]}=\frac{\Pr[M=m1]}{\Pr[M=m1]}$$

$$\frac{\Pr[M=m0 \mid C=c]}{\Pr[M=m0]}=\frac{\Pr[M=m1 \mid C=c]}{\Pr[M=m1]}$$

$$\frac{\Pr[C=c \wedge M=m0]}{\Pr[M=m0]\Pr[C=c]}=\frac{\Pr[C=c \wedge M=m1]}{\Pr[M=m1]\Pr[C=c]}$$

$$\frac{\Pr[C=c \wedge M=m0]}{\Pr[M=m0]}=\frac{\Pr[C=c \wedge M=m1]}{\Pr[M=m1]}$$

$$\Pr[C=c \mid M=m0]=\Pr[C=c \mid M=m1]$$

必要性

已知$\forall i, \Pr[C=c \mid M=mi]=\alpha$

$$\sum_i \Pr[M=mi]=1$$

$$\Pr[C=c]$$

$$=\sum_i \left(\Pr[C=c \mid M=mi]\Pr[M=mi]\right)$$

$$=\alpha \sum_i \Pr[M=mi]$$

$$=\alpha=\Pr[C=c \mid M=mi]$$

敌手能力

唯密文攻击，已知明文攻击，选择明文攻击，选择密文攻击

唯密文攻击实验：

1. 运行密钥生成算法Gen，确定密钥k
2. 将加密任意的明文后发送给攻击者（可反复运行任意多次）
3. 攻击者发送明文空间中的两条明文m0和m1
4. 随机加密其中一条明文，并将对应密文发送给攻击者
5. 将加密任意的明文后发送给攻击者（可反复运行任意多次）
6. 攻击者根据密文猜测对应的明文是m0还是m1

安全目标

唯密文攻击实验：

1. 运行密钥生成算法Gen，确定密钥k

2.将加密任意的明文后发送给攻击者（可反复运行任意多次)

3. 攻击者发送明文空间中的两条明文m0和m1

4. 随机加密其中一条明文，并将对应密文发送给攻击者

5.将加密任意的明文后发送给攻击者（可反复运行任意多次)

6. 攻击者根据密文猜测对应的明文是m0还是m1

定理：

如果加密方案在唯密文攻击下是完善保密的，那么攻击者猜对的概率是1/2。

“一次一密”密码体制（one-time pad）

一次一密加密体制：设$n \geq 1$，$\mathcal{P}, \mathcal{C}, \mathcal{K} = (\mathbb{Z}_2)^n$，对于

$k = (k_1, k_2, \ldots, k_n) \in (\mathbb{Z}_2)^n$，定义

$$e_k(x) = e_k((x_1, x_2, \ldots, x_n)) = (x_1 + k_1, x_2 + k_2, \ldots, x_2 + k_n)$$

$$d_k(y) = d_k((y_1, y_2, \ldots, y_n)) = (y_1 + k_1, y_2 + k_2, \ldots, y_2 + k_n)$$

缺点：

密钥长度和明文长度一样长

密钥只能使用一次

密钥管理异常复杂

机器学习中的PAC理论（Probably Approximately Correct）

一次一密加密体制：设$n \geq 1$，$\mathcal{P}, \mathcal{C}, \mathcal{K} = (\mathbb{Z}_2)^n$，对于

$k = (k_1, k_2, \ldots, k_n) \in (\mathbb{Z}_2)^n$,定义

$$e_k(x) = e_k((x_1, x_2, \ldots, x_n)) = (x_1 + k_1, x_2 + k_2, \ldots, x_2 + k_n)$$

$$d_k(y) = d_k((y_1, y_2, \ldots, y_n)) = (y_1 + k_1, y_2 + k_2, \ldots, y_2 + k_n)$$

从AI角度，给足够多的（密文，明文）做训练集，能否找到一个算法，近似正确地把测试集中的密文映射到明文?

1. 密钥长度与明文长度等长，且明文可以任意长（infinite）
2. 密钥的每位都是从{0,1}中随机决定，所以密钥服从均匀分布
3. 假设集（hypothesis class）可以散布（shatter）在任意大小的样本集合上
4. 所以，假设集的VC维是无限的（infinite），因此不是PAC可学习的