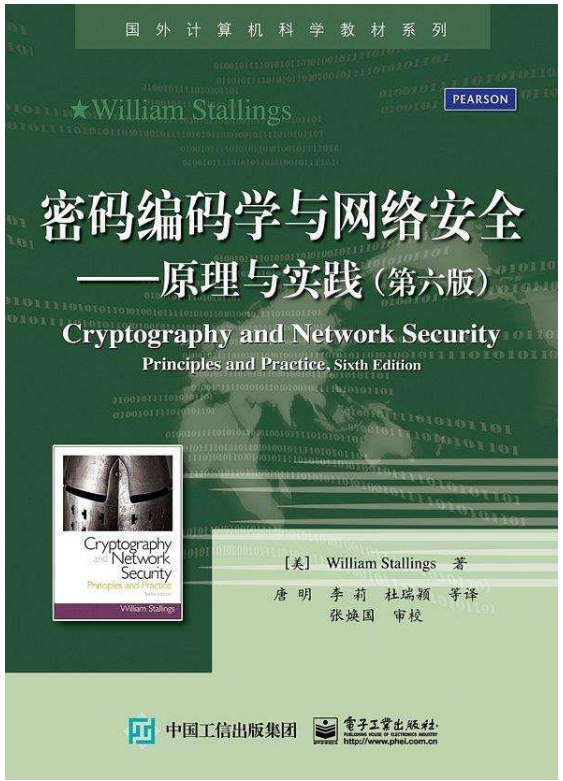
密 码 学

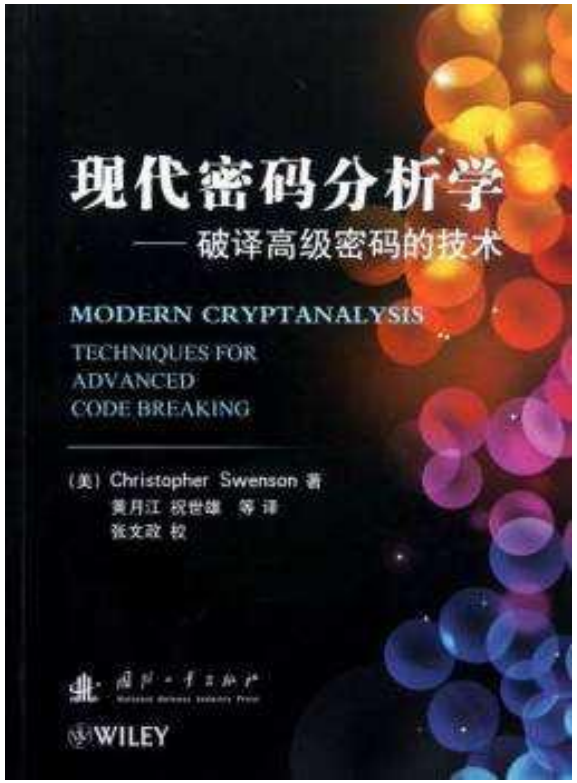
第二章 密码学的基本概念

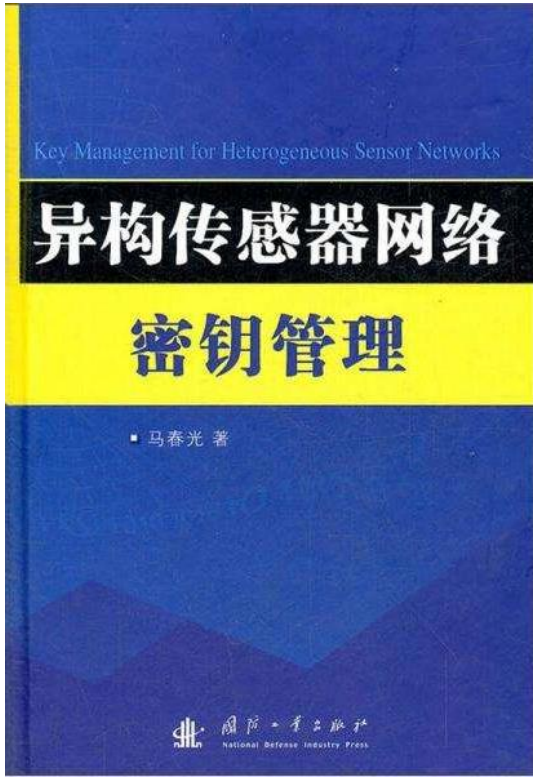
网络空间安全学院
朱 丹
zhudan@nwpu.edu.cn

密码学的内涵

Scope







密码学解决的问题

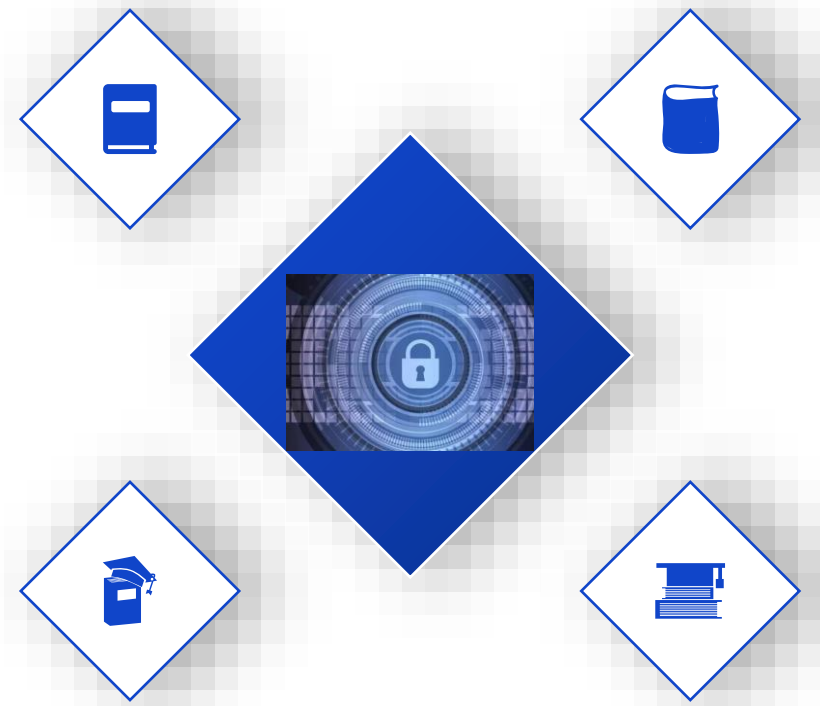
The role of cryptography

机　密　性

防止敏感信息泄漏

完　整　性

防止关键信息被篡改



真　实　性

防止身份或数据假冒

不 可 否 认 性

防止攻击行为抵赖

凡是有机密性、真实性、完整性、不可否认性安全需求的，
都可以用密码技术解决

加密通信模型

Model of Secure Communication

非法接入者

c'

m'

密码分析员

信源

m

加密器

$c = E(k1, m)$

c

公开信道

解密器

$m = D(k2, c)$

m

信宿

密钥源

$k1$

安全信道

密钥源

$k2$

密码学的发展 – 阶段划分

Developments - Stages

1

古 典 密 码

从古代到19世纪末 - 凯撒密码、维吉尼亚密码

2

近 代 密 码

从20世纪初到1949年 - 恩尼格玛(Enigma)密码机

3

现 代 密 码

从1949年到1976年 – 序列密码、DES

4

公 钥 密 码

从1976年开始 – RSA、ECC

5

下 一 阶 段 ？

后量子密码

密码学的发展 – 第二阶段

Developments - Stages

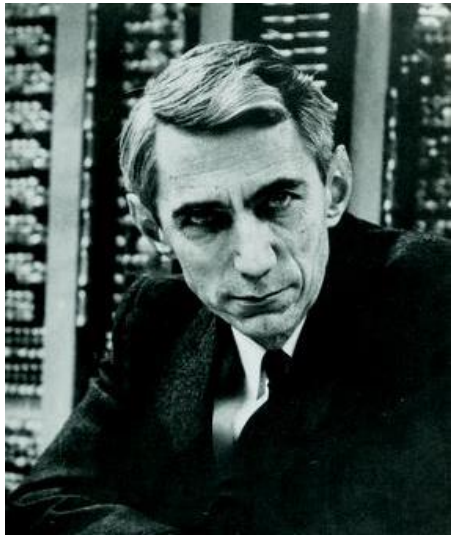
对称密码学早起发展时期（1949– 1975）

1949年，Shannon发表题为《保密系统的通信理论》（Communication Theory of Secrecy Systems）的论文

该文为对称密码系统建立了理论模型

该文创立的信息论为密码学奠定了理论基础

从此，密码学发展成为一门真正的科学



Claude E. Shannon

密码学的发展 – 第三阶段

Developments - Stages

现代密码学发展时期（1976 – 1996）

1976年，Diffie和Hellman发表题为《密码学的新方向》（New Directions in Cryptography）的论文

该文引入了公钥密码的概念

1977年，美国制定了数据加密标准（Data Encryption Standard，DES）

公钥密码和DES标志着现代密码学的诞生

美国计算机协会（ACM）将2015年的图灵奖授予Sun Microsystems的前首席安全官惠特菲尔德·迪菲（Whitfield Diffie）以及斯坦福大学电气工程系名誉教授马丁·赫尔曼（Martin Hellman），以表彰他们在现代密码学中所起的至关重要的作用。

密 码 算 法 的 要 求

Considerations

不可破译性：理论和实际上都是不可破译的

算法覆盖性：覆盖整个密钥空间

一切秘密寓于密钥之中（Kerckhoffs原则）

实现性能：便于实现，性能好

奥古斯特·柯克霍夫在19世纪提出：密码系统应该就算被所有人知道系统的运作步骤，仍然是安全的。

克劳德·艾尔伍德·香农有句近似的话「敌人知道系统」，称为香农公理。

Kerckhoffs

密码体制安全性

Security of Crypto System

无　条　件　安　全

密码分析者具有无限的计算能力，密码体制也不能破译

3

计　算　安　全

当前计算条件（时间和存储器资源）无法满足密码破译的需求

2

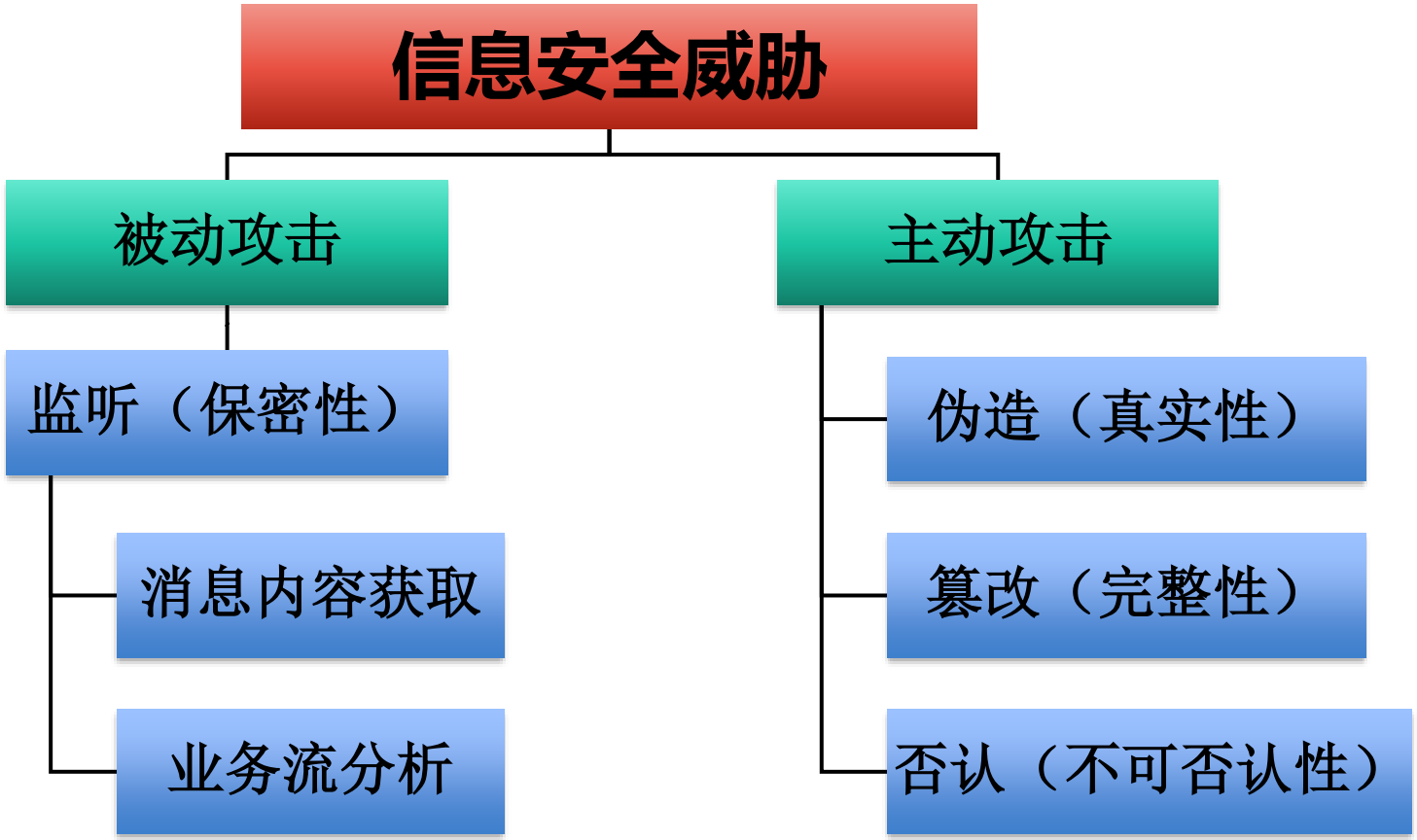
可　证　明　安　全

把密码体制的安全性规约位某个经过深入研究的数学难题

1

信息安全威胁

Information security threats

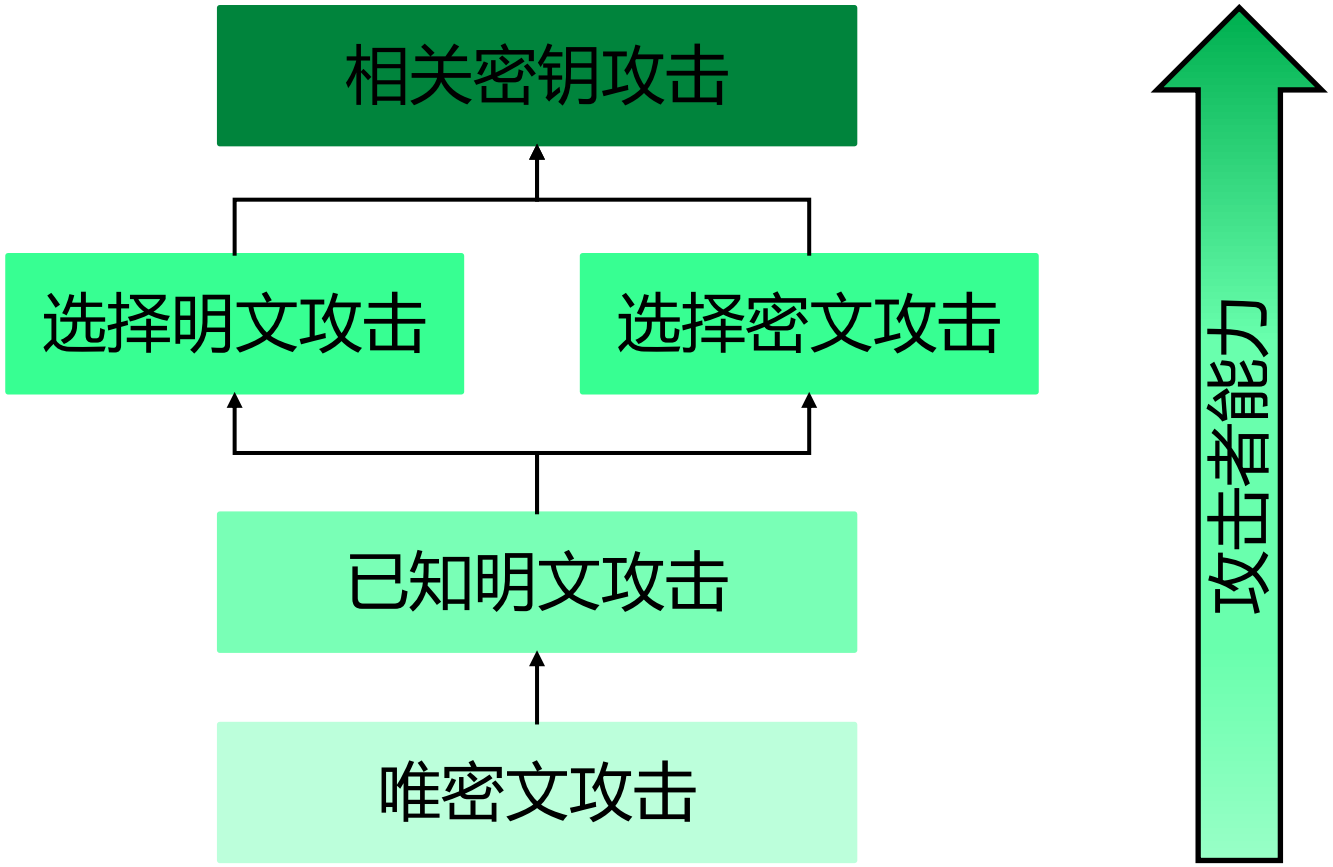


密 码 分 析 学

Cryptanalysis

主要任务：破译密码或伪造消息

按照攻击者掌握的信息类别（攻击者能力）分为：



章节安排

Outline

密码学的基本概念

密码学发展史

密码体制的安全性

密码编码的基本方法

替代密码的统计分析

章节安排

Outline

密码学的基本概念

密码学发展史

密码体制的安全性

密码编码的基本方法

替代密码的统计分析

置 换 与 替 代

Permutation and Substitution

01 置 换 密 码

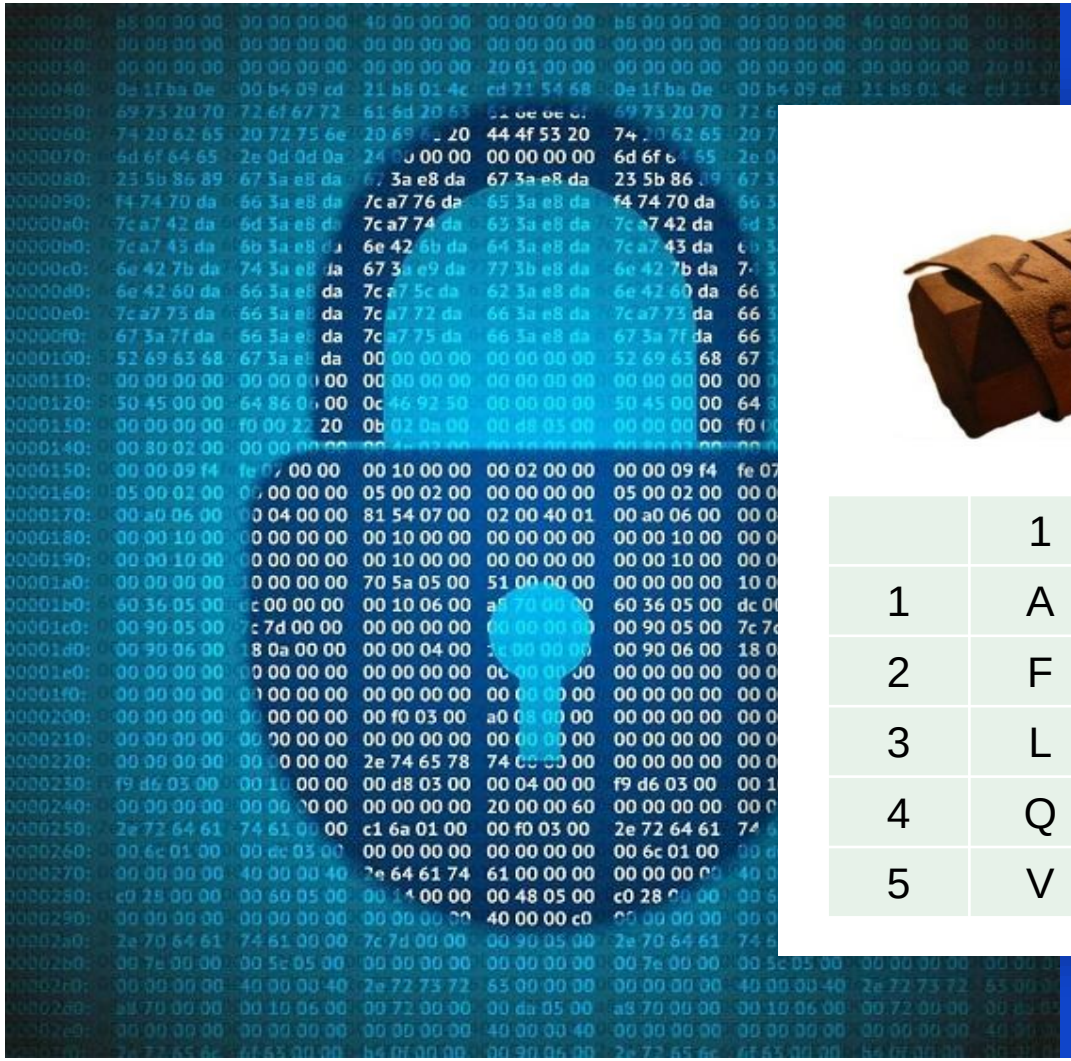
置换密码算法的原理是不改变明文字符，只将字符在明文中的排列顺序改变，从而实现明文信息的加密。置换密码有时又称为换位密码

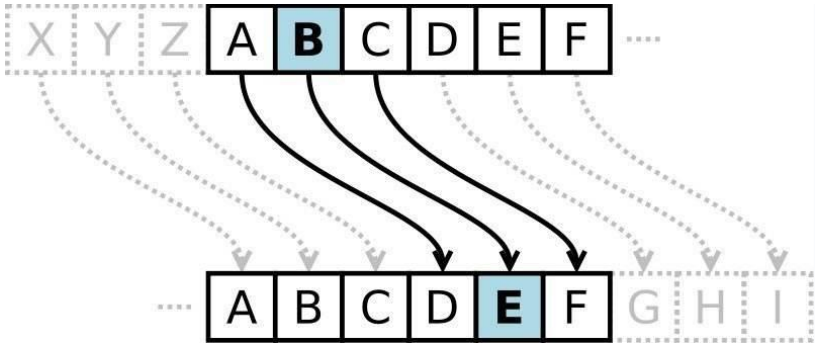
02 替 代 密 码

替代密码替代密码算法的原理是使用替代法进行加密，就是将明文中的字符用其它字符替代后形成密文。加密后明文字符的形态会发生变化

置 换 与 替 代

Permutation and Substitution





	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	IJ	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

置 换 密 码

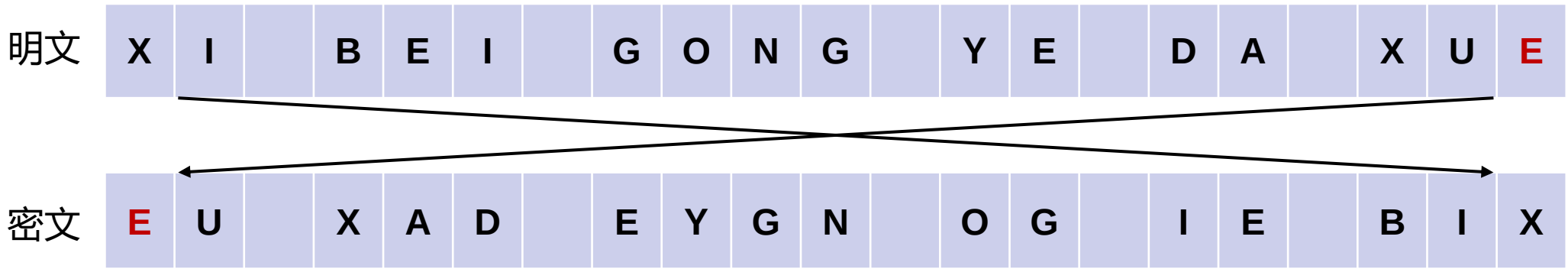
Permutation Cipher

是对明文字符的位置重新进行排列的一种密码

亦称为换位密码

逆序选出

例1.1，明文：西北工业大学



置换方式：把明文中的字母顺序反过来

置 换 密 码

Permutation Cipher

1	2	3	4	5	6
X	I	B	E	I	G
O	N	G	Y	E	D
A	X	U	E		

是对明文字符的位置重新进行排列的一种密码

亦称为换位密码

按行重写，按列选出

例1.2，明文：西北工业大学																			
明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E
密文	X	O		A	I	N		X	B	G	U		E	Y		E	I		E	G	D

置换方式：把明文按照某一顺序排列成矩阵，然后以列为顺序选出矩阵中的字母作为密文

置 换 密 码

Permutation Cipher

数字顺序	7	1	4	2	3	6	5
	W	A	N	G	L	U	O
选出顺序	X	I	B	E	I	G	O
	N	G	Y	E	D	A	X
	U	E					

是对明文字符的位置重新进行排列的一种密码

亦称为换位密码

以密钥字母数字顺序按列选出

例1.3，明文：西北工业大学

明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E

密 钥 ： 网 络　W A N G　L U O → W A N G L U O → 7 1 4 2 3 6 5

密文	I	G		E	E	E		I	D	B	Y		O	X		G	A		X	N	U

置换方式：把明文按照某一顺序排列成矩阵，然后按照密钥字母顺序选出矩阵中的字母作为密文

置 换 密 码

Permutation Cipher

置换密码的特点

打乱了明文字母之间的跟随关系，使得明文自身的结构规律也得到破坏

置换密码的缺点

明文字符的形态不变（X → X）

密文字符出现的频次与对应的明文字符出现的频次相同

简单的唯密文攻击或已知明文攻击即可破译置换密码

替 代 密 码

Substitution Cipher

替代密码是利用预先定义的代替规则，对明文逐字符进行替代的密码算法

替代密码的替代规则就是其密钥

替代规则又称为替代函数、替代表或S盒

单 表 替 代 密 码

只使用一个密文字符表

多 表 替 代 密 码

使用多组密文字符表或映射函数

单 表 替 代 密 码

Substitution Cipher

单表替代密码又称单替代密码

只使用一个密文字符表

用密文字母表中的一个字符对应替代明文的一个字符

设A和B分别为含n个字符的明文字符表和密文字符表：

$$A = \{a_0, a_1, a_2, \ldots, a_{n-1}\}$$

$$B = \{b_0, b_1, b_2, \ldots, b_{n-1}\}$$

定义一个由A到B的一一映射$f : A \rightarrow B$

$$f(a_i) = b_i$$

设明文$M = \{m_0, m_1, m_2, \ldots, m_{n-1}\}$，则密文$C = \{f(m_0), f(m_1), f(m_2), \ldots, f(m_{n-1})\}$

单表替代密码的密钥就是映射函数 f 或密文字符表B

单表替代密码实例 – 加法密码

Substitution Cipher

加法密码的映射函数为

$$f(a_i) = b_i = a_j$$

$$j = i + k \bmod n$$

其中， $a_i \in A$, k是满足$0 < k < n$的正整数， $0 \le i, j < n$

设A和B分别为含n个字符的明文字符表和密文字符表：

$$A = \{a_0, a_1, a_2, \ldots, a_{n-1}\}$$

$$B = \{b_0, b_1, b_2, \ldots, b_{n-1}\}$$

定义一个由A到B的——映射f : A → B

$$f(a_i) = b_i$$

设明文$M = \{m_0, m_1, m_2, \ldots, m_{n-1}\}$，则密文$C = \{f(m_0), f(m_1), f(m_2), \ldots, f(m_{n-1})\}$

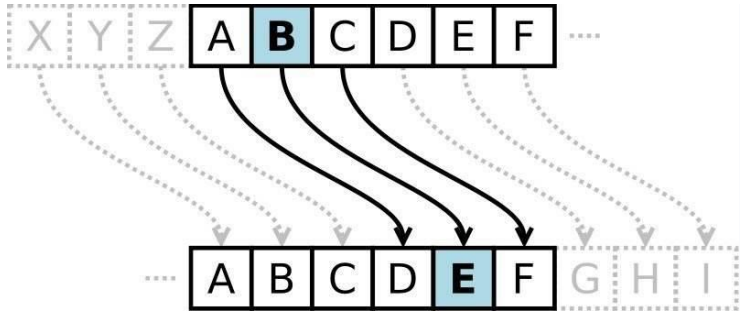
单表替代密码实例 – 凯撒密码

Substitution Cipher

凯撒（Caesar）密码即是著名的加法密码

凯撒密码取$k = 3$

$$j = i + 3 \bmod 26$$



根据偏移量的不同，还存在若干特定的恺撒密码名称：

- 偏移量为10：Avocat(A→K)
- 偏移量为13：ROT13
- 偏移量为-5：Cassis (K 6)
- 偏移量为-6：Cassette (K 7)

例1.4，明文：西北工业大学

明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E
密文	A	L		E	H	L		J	R	Q	J		B	H		G	D		A	X	H

置换方式：密文字符由明文字符循环右移3位得到

单表替代密码实例 – 乘法密码

Substitution Cipher

乘法密码的映射函数为

$$f(a_i) = b_i = a_j$$
$$j = i * k \bmod n,\ 0 \le i, j < n,\ (k, n) = 1$$

当用英文字母表作为明文字母表时（$n = 26$），若取$k = 13$，密文无法实现解密

$$f(A) = f(C) = f(E) = \ldots = f(Y) = A$$
$$f(B) = f(D) = f(F) = \ldots = f(Z) = N$$

k要与n互素时，才存在两个整数x，y使得$xk + yn = 1$，才有$xk = 1 \bmod n$，

进而有$xj \bmod n = x\ (i * k \bmod n) \bmod n = i\ (xk \bmod n) \bmod n = i$

单表替代密码实例 – 乘法密码

Substitution Cipher

若取$k = 5$ ($n = 26$)，由$j = i * 5 \bmod 26$，可得到如下的密文字符表

$$A = \{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z\}$$

$$B = \{A, F, K, P, U, Z, E, J, O, T, Y, D, I, N, S, X, C, H, M, R, W, B, G, L, Q, V\}$$

例1.5，明文：西北工业大学

明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E
密文	L	O		F	U	O		E	S	N	E		Q	U		P	A		L	W	U

置换方式：根据规则计算或查表

单表替代密码实例 – 仿射密码

Substitution Cipher

加法密码和乘法密码相结合便构成仿射密码

$$f(a_i) = b_i = a_j$$

$$j = i * k_1 + k_0 \bmod n$$

其中，$(k_1, n) = 1$，且$0 < k_0 < n$。

进一步可构造更复杂的多项式密码

$$f(a_i) = b_i = a_j$$

$$j = i^t * k_t + i^{t-1} * k_{t-1} + \ldots i * k_1 + k_0 \bmod n$$

其中，k_i要与n互素，即 $(k_i, n) = 1$ $(i = 1, 2, \ldots, t)$，且$0 < k_0 < n$。

多 表 替 代 密 码

Substitution Cipher

基本思想

单表替代密码的信息泄露本质上都是由于一个明文字符总是被一个固定的密文字符替代所造成的

如果一个明文字符可能被多个密文字符替代，那么密文字符组成的密文字符串的统计规律就可能变得均匀，从而更安全

构造d个密文字符表

$$B_j = \{b_{j0}, b_{j1}, \cdots, b_{jn-1}\}, \quad j = 0, 1, 2, \cdots, d-1$$

定 义 d 个 映 射

$$f_j: A \to B_j$$
$$f_j(a_i) = b_{ji}$$

设明文$M = (m_0, m_1, m_2, \ldots, m_{d-1}, m_d, \ldots)$，则密文为

$$C = (f_0(m_0), f_1(m_1), f_2(m_2), \ldots, f_{d-1}(m_{d-1}), f_d(m_d, \ldots))$$

多表替代密码实例 – Vigenere密码

Substitution Cipher

16世纪法国密码学者Vigenere使用过的

Vigenere密码是一种著名的多表替代密码

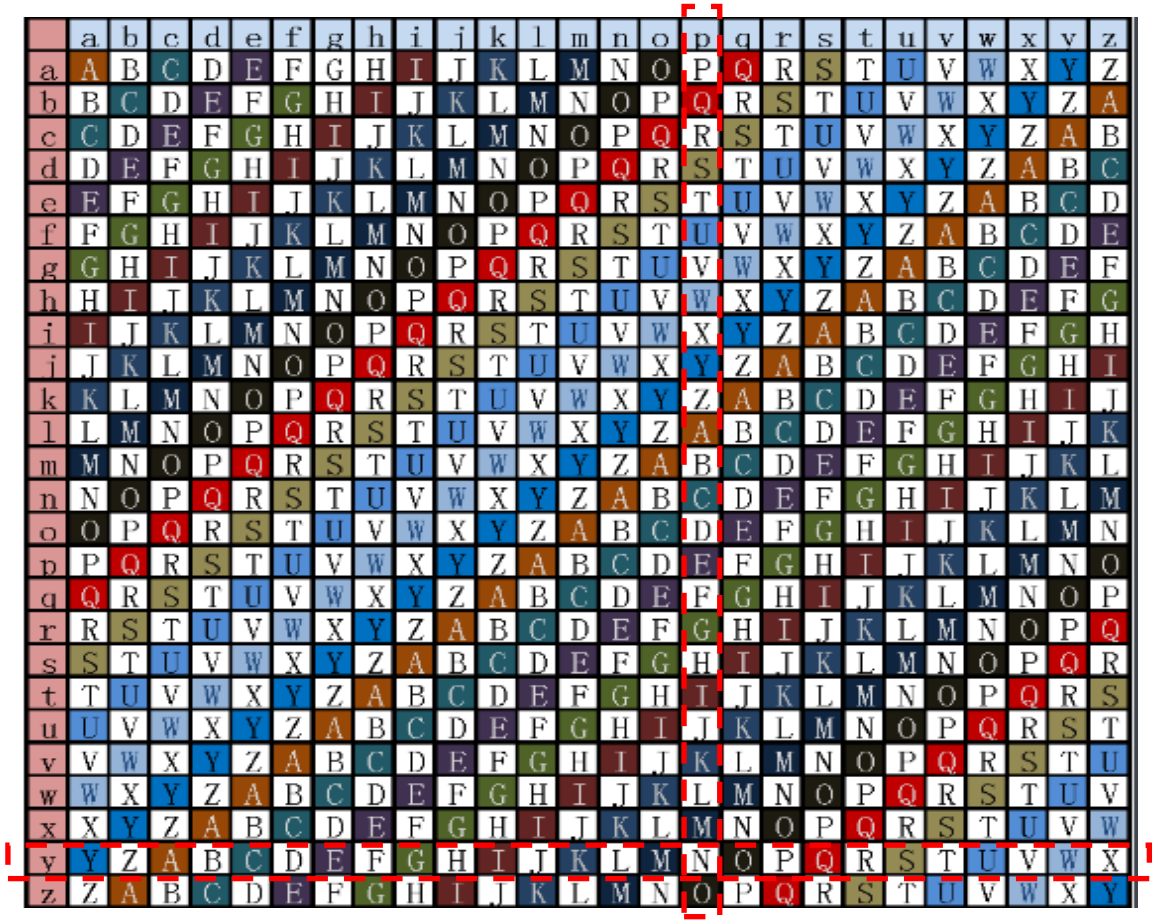
26个密文字符表

明文字符表循环移位1-25位

选用一个短语作为密码，密钥字符用于选择密文字符替代表

例如，明文字符为P，密钥字符为Y

密文字符为N



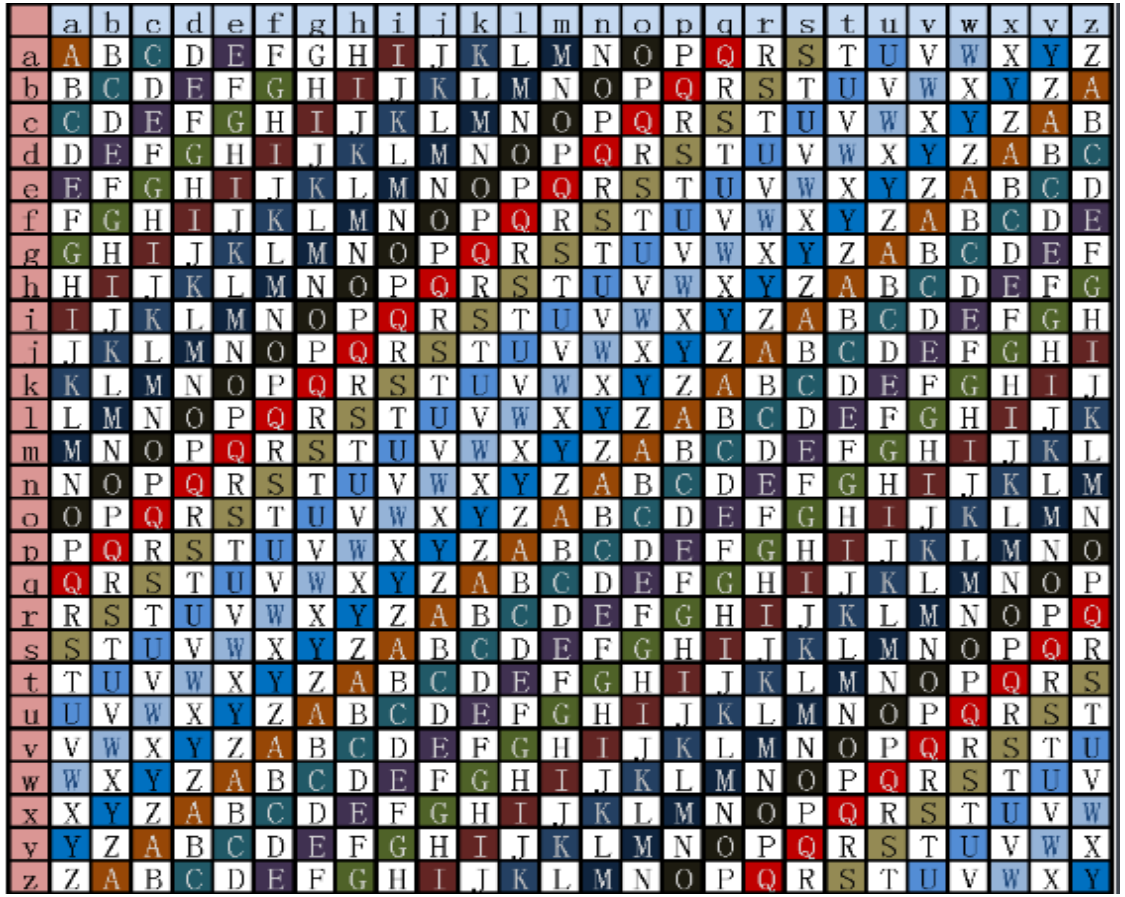
多表替代密码实例 – Vigenere密码

Substitution Cipher

选用一个短语作为密码，密钥字符用于选择密文字符替代表

例如，明文字符为B，密钥字符为N

密文字符为O



明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E
密钥	W	A		N	G	L		U	O	A	N		Q	U		A	N		Z	A	O
密文	T	I		O	K	T		A	C	N	T		O	Y		D	N		W	U	S

多表替代密码实例 – Vernam密码

Substitution Cipher

Gillbert Vernam于1917年为电报通信设计的一种非常方便的密码

Vernam密码奠定了序列密码的基础

Vernam密码的明文、密钥和密文都为二进制序列

设明文$M = (m_0, m_1, \ldots, m_{n-1})$，密钥$K = (k_0, k_1, \ldots, k_{n-1})$，密文$C = (c_0, c_1, \ldots, c_{n-1})$，则

$$c_i = m_i \oplus k_i$$
$$m_i = c_i \oplus k_i$$
$$i = 0, 1, \ldots, n\text{-}1$$

多表替代密码实例 – Vernam密码

Substitution Cipher

Vernam密码举例:

明文：DATA

密钥：LAMB

明文	1 0 0 0 1 0 0	1 0 0 0 0 0 1	1 0 1 0 1 0 0	1 0 0 0 0 0 1
密钥	1 0 0 1 1 0 0	1 0 0 0 0 0 1	1 0 0 1 1 0 1	1 0 0 0 0 1 0
密文	0 0 0 1 0 0 0	0 0 0 0 0 0 0	0 0 1 1 0 0 1	0 0 0 0 0 1 1

多表替代密码的特点

Substitution Cipher

多表替代密码的密钥就是这组映射函数或密文字符表

明文中相同的字符不再总是被映射成相同的字符

明文的统计规律不在反映在密文中

如果密钥序列是随机的，即他们相互独立且服从均匀分布，则在未知密钥序列的条件下，即使知道加密变换，该密码也是不可破译的

总　结

Summary

01 替代密码和置换密码相结合

在实际应用中，通常将替代密码和置换密码结合起来，从而设计出安全的密码体制

02 替　代　-　置　换　模　型

这就是分组密码的替代-置换模型，即现代密码的设计思想：利用简单的密码变换的组合，设计出抗攻击能力强的密码算法

章节安排

Outline

密码学的基本概念

密码学发展史

密码体制的安全性

密码编码的基本方法

替代密码的统计分析

单表替代密码的统计分析

Cryptanalysis

加法密码的破译

k只有n-1种取值（$0 < k < n$）

明文为字母表时，$n = 26$，k只有25种取值

即使明文为8位扩展ASCII码，$n = 256$，k只有255种取值

穷举法破解

$$f(a_i) = b_i = a_j$$

$$j = i + k \bmod n$$

其中，$a_i \in A$，k是满足$0 < k < n$的正整数

单表替代密码的统计分析

Cryptanalysis

乘法密码的破译

加法密码和乘法密码哪个相对更容易破译?

单表替代密码的统计分析

Cryptanalysis

乘法密码的破译

还要求（k, n）= 1，比加法密码的密钥空间更小

明文为字母表时，$n = 26$，k的取值只能是3，5，7，9，11，15，17，19，21，23，25这11种

即使明文为8位扩展ASCII码，$n = 256$，k有多少种取值?

穷举法破解

$$f(a_i) = b_i = a_j$$

$$j = i * k \bmod n$$

其中，k要与n互素，仅当（k, n）= 1时，才能正确解密

单表替代密码的统计分析

Cryptanalysis

仿射密码的破译

比加法密码和乘法密钥的保密性要好一些

密钥也只有$n * (\varphi(n) - 1)$种，其中$\varphi(n)$为欧拉函数

明文为字母表时，$n = 26$，可能的密钥只有$26 * (12 - 1) = 286$种

穷举法破解

$$f(a_i) = b_i = a_j$$

$$j = i * k_1 + k_0 \bmod n$$

其中，k_1要与n互素，即$(k, n) = 1$，且$0 < k_0 < n$。

英文字母频率分布

Statistics of English Characters

字母	A	B	C	D	E	F	G	H	I	J	K	L	M
频率	8.167	1.492	2.782	4.253	12.702	2.228	2.015	6.094	6.966	0.153	0.722	4.025	2.406
字母	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
频率	6.749	7.507	1.929	0.095	5.987	6.327	9.056	2.758	0.978	2.360	0.150	1.974	0.074

Londoners are under starter's orders as the city gets ready for the Olympic Games, which will begin one year today. To mark the start of the 366-day countdown (2012 is a leap year), special events are planned for today. The design of the Olympic medals will be unveiled tonight in a live ceremony from Trafalgar Square. Over at the brand new Aquatics Centre, Britain's star diver Tom Daley is going to perform an official launch dive into the Olympic pool. With this building, the organizers have attempted to give London a landmark to rival Beijing's Water Cube from 2008. It was designed by the prestigious architect Zaha Hadid and has a wave-like roof that is 160 meters long. Today's special events are designed to arouse interest in the Olympics around the world and to encourage British fans too. Many failed to get Olympic tickets in the recent sales process. According to a new survey for the BBC, 53% of Londoners think the process was "not fair". But the same survey found support is growing for London 2012. Of the 1,000 people surveyed, 73% said they backed the Games - up from 69% in 2006. Olympics minister Hugh Robertson said: "We are under budget and ahead of time and as a nation we have a reputation of really getting behind these big events."

英文字母频率分布

Statistics of English Characters

单字母频率

极高频率字母	E
次高频率字母组	T A O I N S H R
中等频率字母组	D L
低频率字母组	C U M W F G Y P B
甚低频率字母组	V K J X Q Z

出现频率最高的30个双字母

TH HE IN ER AN RE ED ON ES ST EN AT TO NT HA ND
OU EA NG AS OR TI IS ET IT AR TE SE HI OF

出现频率最高的20个三字母

THE ING AND HER ERE ENT THA NTH WAS ETH FOR
DTH HAT SHE ION HIS STH ERS VER

英文字母频率分布

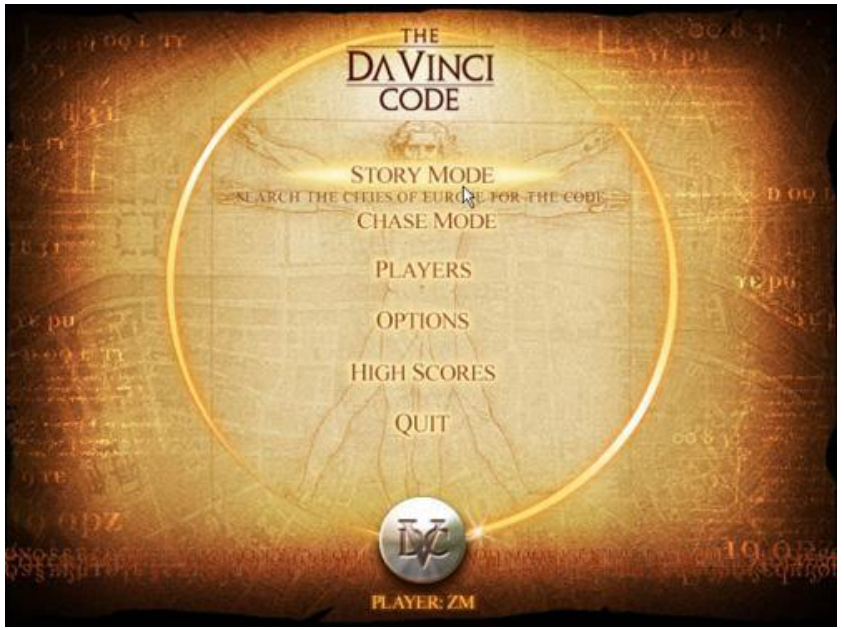
Statistics of English Characters

其他统计规律

英文单词以E、S、D、T为结尾的约占一半

英文单词以T、A、S、W为起始的字母约占一半

密码分析者的文学、历史、地理等方面的知识对于密码破译也是十分重要的因素



单 表 替 代 密 码 的 统 计 分 析

Cryptanalysis

单表替代密码统计分析

首先，统计密文的各种统计特征

其次，分析双字母、三字母密文组，以区分元音和辅音字母

最后，分析字母较多的密文，可以使用猜测法

字母	A	B	C	D	E	F	G	H	I	J	K	L	M
频率	8.167	1.492	2.782	4.253	12.702	2.228	2.015	6.094	6.966	0.153	0.722	4.025	2.406
字母	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
频率	6.749	7.507	1.929	0.095	5.987	6.327	9.056	2.758	0.978	2.360	0.150	1.974	0.074

多表替代密码的统计分析

Cryptanalysis

多表替代密码的密文字符与明文字符相比，其频率分布更趋于平均

假设多表替代密码的周期为d，明文字符中每个字符将可能被*d*种不同的字符替代

明文	X	I		B	E	I		G	O	N	G		Y	E		D	A		X	U	E
密钥	W	A		N	G	L		U	O	A	N		Q	U		A	N		Z	A	O
密文	T	I		O	K	T		A	C	N	T		O	Y		D	N		W	U	S

为定量分析周期多替代密码的频率分布于单表替代密码的频率分布的区别，引入粗糙度和重合指数的概念

粗　　糙　　度

Measure of Roughness

粗糙度(Measure of Roughness, M.R)定义为每个密文字母出现的频率与均匀分布时每个字母出现的频率之差的平方和

设各密文字母出现的频率为p_i $(i = 0, 1, 2, \ldots, 25)$，则有$\sum_{i=0}^{25} p_i = 1$

对于英文报文，则$n = 26$。均匀分布下，每个字母出现的概率为1/26

$$M.R = \sum_{i=0}^{25} (p_i - \frac{1}{26})^2$$

$$= \sum_{i=0}^{25} p_i^2 - \frac{1}{26} = \sum_{i=0}^{25} p_i^2 - 0.0385$$

粗　　糙　　度

Measure of Roughness

单表替代密码密文字母频率分布表

字母	A	B	C	D	E	F	G	H	I	J	K	L	M
频率	8.167	1.492	2.782	4.253	12.702	2.228	2.015	6.094	6.966	0.153	0.722	4.025	2.406
字母	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
频率	6.749	7.507	1.929	0.095	5.987	6.327	9.056	2.758	0.978	2.360	0.150	1.974	0.074

$$\sum_{i=0}^{25} p_i^2 - 0.0385 = 0.0655 - 0.0385 = 0.027$$

由此，单表代替或明文的粗糙度约为0.027，若字符均匀分布，则报文的粗糙度为0。因此，粗糙度一般在0～0.027之间变化

如计算出密文段的粗糙度为0.006366，则可断定，该密文段采用多表代替密码加密得到

重　合　指　数

Index of Coincidence

重合指数（Index of Coincidence, IC）的概念由Friedman于1918年提出，其论文《重合指数及其在密码学中的应用》是1949年以前最有影响的密码学文献

William F. Friedman

定义：设某种语言由n个字母组成，第i个字母出现的概率为p_i，$0 \leq i < n$，重合指数是指两个随机字母相同的概率：

$$IC = \sum_{i=0}^{n-1} p_i^2$$

重　合　指　数

Index of Coincidence

单表代替情况下，明文和密文的IC是相同的（对于英文，均为0.0655），即单表替代不改变频率分布

多表代替情况下，密文的IC值较小，密文趋向均匀分布

IC值可用来判断密文采用了多表代替加密还是单表代替加密

可通过计算IC值，看其是否接近0.0655，来分析多表代替的密钥长度

课 后 作 业

Homework

作业2.1：利用仿射密码，选择合适的加密算法参数，对以下明文内容（见plaintext.txt）进行加密，给出加密结果第一段

作业2.2：利用所学方法对作业2.1中的完整加密结果进行解密，给出详细的密码分析过程


感谢聆听!
THANK YOU FOR YOUR ATTENTION!