一些简单密码体制的破译

杜育松

中山大学计算机学院

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

移位密码

密码体制1 (移位密码)

设$\mathcal{P}=\mathcal{C}=\mathcal{K}=\mathbb{Z}_{26}$。对于$0\leq K\leq 25$和任意的$x,y\in\mathbb{Z}_{26}$，加密规则定义为

$$e_K(x)=(x+K)\bmod 26.$$

解密规则定义为

$$d_K(y)=(y-K)\bmod 26.$$

特别地，如果$K=3$，则此密码体制通常被称为凯撒密码。

代换密码

密码体制2 (代换密码)

设$\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$。密钥空间$\mathcal{K}$是由26个数字$0, 1, 2 \cdots, 25$的所有可能的置换组成。对于置换$\pi \in \mathcal{K}$和任意的$x, y \in \mathbb{Z}_{26}$，定义加密规则为

$$e_\pi(x) = \pi(x).$$

解密规则为

$$d_\pi(y) = \pi^{-1}(y).$$

仿射密码

密码体制3 (仿射密码)

设$\mathcal{P}=\mathcal{C}=\mathbb{Z}_{26}$。密钥空间$\mathcal{K}=\{(a,b)\in\mathbb{Z}_{26}\times\mathbb{Z}_{26}:\gcd(a,26)=1\}$。对于密钥$K=(a,b)\in\mathcal{K}$和任意的$x,y\in\mathbb{Z}_{26}$，定义加密规则为

$$e_K(x)=(ax+b)\bmod 26,$$

解密规则定义为

$$d_K(y)=a^{-1}(y-b)\bmod 26.$$

- 仿射密码可以解密成功因为有同余式

$$d_K(y)\equiv d_K((ax+b))\equiv a^{-1}(ax+b-b)\equiv a^{-1}ax\equiv x\bmod 26$$

成立。

- 仿射密码所有能的密钥一共有多少个呢？答案是$26\times 12=312$。

维吉尼亚密码

密码体制4 (维吉尼亚密码)

设$\mathcal{P}=\mathcal{C}=\mathcal{K}=(\mathbb{Z}_{26})^m$。对于密钥$K=(k_1,k_2,\cdots,k_m)$和任意的

$$x=(x_1,x_2,\cdots,x_m)\in(\mathbb{Z}_{26})^m, y=(y_1,y_2,\cdots,y_m)\in(\mathbb{Z}_{26})^m,$$

定义加密规则为

$$e_K(x_1,x_2,\cdots,x_m)=(x_1+k_1,x_2+k_2,\cdots,x_m+k_m),$$

解密规则定义为

$$d_K(y_1,y_2,\cdots,y_m)=(y_1-k_1,y_2-k_2,\cdots,y_m-k_m),$$

并且以上运算都在$\mathbb{Z}_{26}$上进行。

- 维吉尼亚密码是m维向量形式的移位密码。

从单表代换到多表代换

定义1 (单表代换)

一旦密钥被选定，每个字母都被加密规则变换成唯一的密文字母，就好像有唯一的一张事先确定好了的表一样。例如，移位密码、代换密码和仿射密码都是单表代换的密码体制。

定义2 (多表代换)

密钥被选定后，相同字母仍有可能被加密规则变换成不同的密文字母，就好像有多张表一样不断地变换使用。维吉尼亚密码是多表代换的密码体制，一个字母可以被映射为m个字母中的某一个，即有m 种可能。

希尔密码

密码体制5 (希尔密码)

设$\mathcal{P}=\mathcal{C}=(Z_{26})^m$。密钥空间$\mathcal{K}$是定义在$\mathbb{Z}_{26}$上的所有$m\times m$可逆矩阵。对于密钥$\mathbf{K}\in\mathcal{K}$和任意的

$$\mathbf{x}=(x_1,x_2,\cdots,x_m)\in(\mathbb{Z}_{26})^m$$

和

$$\mathbf{y}=(y_1,y_2,\cdots,y_m)\in(\mathbb{Z}_{26})^m,$$

定义加密规则为

$$e_K(\mathbf{x})=\mathbf{x}\cdot\mathbf{K}\,\mathrm{mod}26,$$

解密规则为

$$d_K(\mathbf{y})=\mathbf{y}\cdot\mathbf{K}^{-1}\,\mathrm{mod}26.$$

置换密码

密码体制6 (置换密码)

设$\mathcal{P}=\mathcal{C}=(Z_{26})^m$，密钥空间$\mathcal{K}$是由$m$个数字$1,2\cdots,m$的所有可能的置换组成。对于置换$\pi\in\mathcal{K}$和任意的$\mathbf{x}=(x_1,x_2,\cdots,x_m)\in(\mathbb{Z}_{26})^m$和$\mathbf{y}=(y_1,y_2,\cdots,y_m)\in(\mathbb{Z}_{26})^m$，定义加密规则为

$$e_\pi(\mathbf{x})=(x_{\pi(1)},x_{\pi(2)},\cdots,x_{\pi(m)}),$$

解密规则为

$$d_\pi(\mathbf{y})=(y_{\pi^{-1}(1)},y_{\pi^{-1}(2)},\cdots,y_{\pi^{-1}(m)}).$$

这里，π^{-1}是π的逆置换。

- 代换密码利用置换的结果加密，而置换密码利用置换的方式加密。

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

密码分析中的Kerckhoff假设

- 密码分析的最基本原则是Kerckhoff假设：假定密码分析者知道加密者使用的密码系统，即不把安全性基于对手不知道所用的密码体制。

密码分析的四种模型

1. 唯密文攻击（Ciphertext-Only Attack）：密码分析者仅仅知道加密算法和待破译的密文。

密码分析的四种模型

1. 唯密文攻击（Ciphertext-Only Attack）：密码分析者仅仅知道加密算法和待破译的密文。
2. 已知明文攻击（Known-Plaintext Attack）：密码分析者不仅知道密码算法，还有一定数量的明文和对应的密文。

密码分析的四种模型

1. 唯密文攻击（Ciphertext-Only Attack）：密码分析者仅仅知道加密算法和待破译的密文。
2. 已知明文攻击（Known-Plaintext Attack）：密码分析者不仅知道密码算法，还有一定数量的明文和对应的密文。
3. 选择明文攻击（Chosen-Plaintext Attack）：密码分析不仅者知道密码算法，还能选择一些明文并获得所对应的密文。

密码分析的四种模型

1. 唯密文攻击（Ciphertext-Only Attack）：密码分析者仅仅知道加密算法和待破译的密文。
2. 已知明文攻击（Known-Plaintext Attack）：密码分析者不仅知道密码算法，还有一定数量的明文和对应的密文。
3. 选择明文攻击（Chosen-Plaintext Attack）：密码分析不仅者知道密码算法，还能选择一些明文并获得所对应的密文。
4. 选择密文攻击（Chosen-Ciphertext Attack）：密码分析不仅者知道密码算法，还能选择一些密文并获得所对应的明文。

密码分析的四种模型

1. 唯密文攻击（Ciphertext-Only Attack）：密码分析者仅仅知道加密算法和待破译的密文。
2. 已知明文攻击（Known-Plaintext Attack）：密码分析者不仅知道密码算法，还有一定数量的明文和对应的密文。
3. 选择明文攻击（Chosen-Plaintext Attack）：密码分析不仅者知道密码算法，还能选择一些明文并获得所对应的密文。
4. 选择密文攻击（Chosen-Ciphertext Attack）：密码分析不仅者知道密码算法，还能选择一些密文并获得所对应的明文。

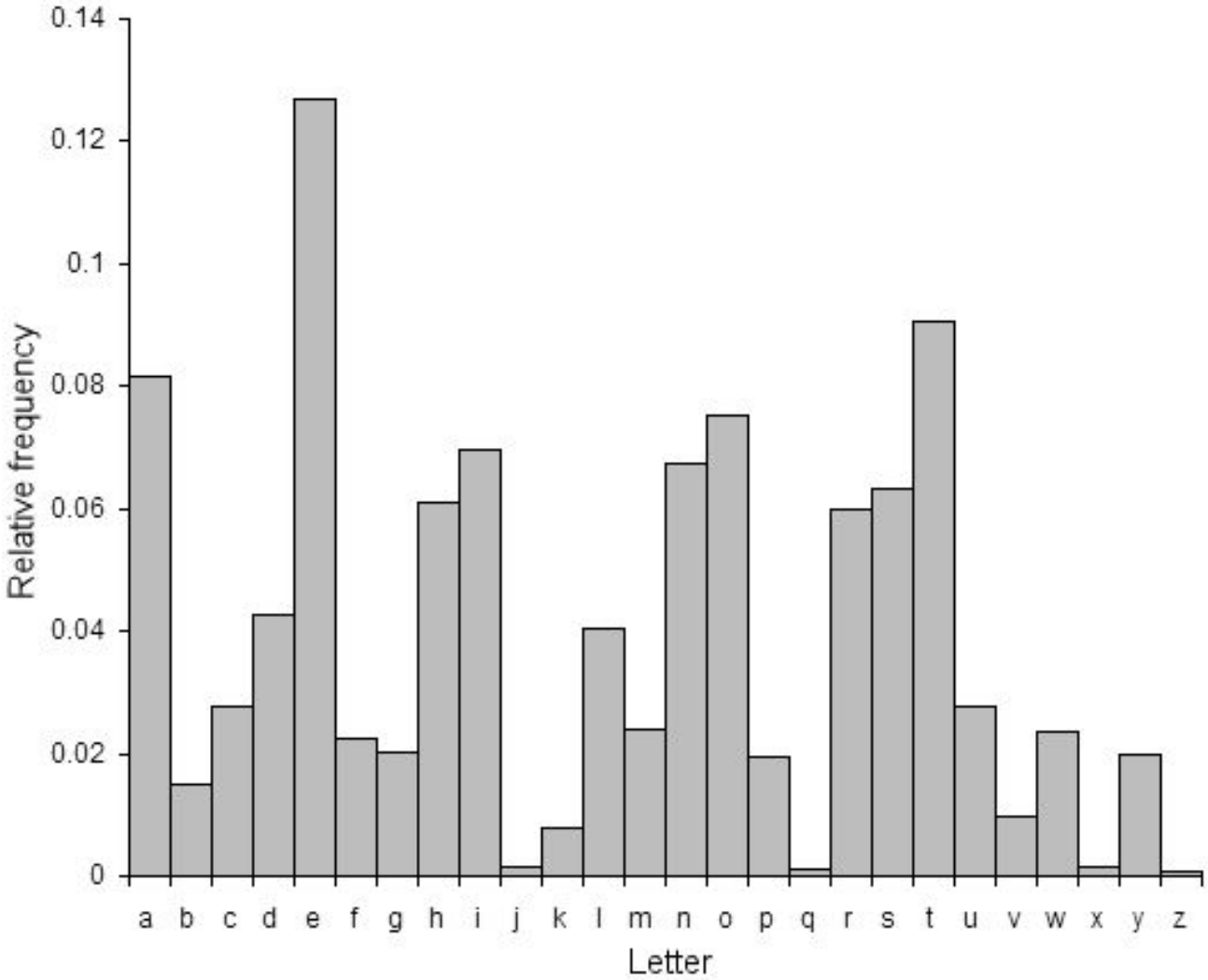
- 在以上四种攻击类型中，选择密文攻击的攻击强度最高，选择明文攻击次之。

英语字母的统计特征

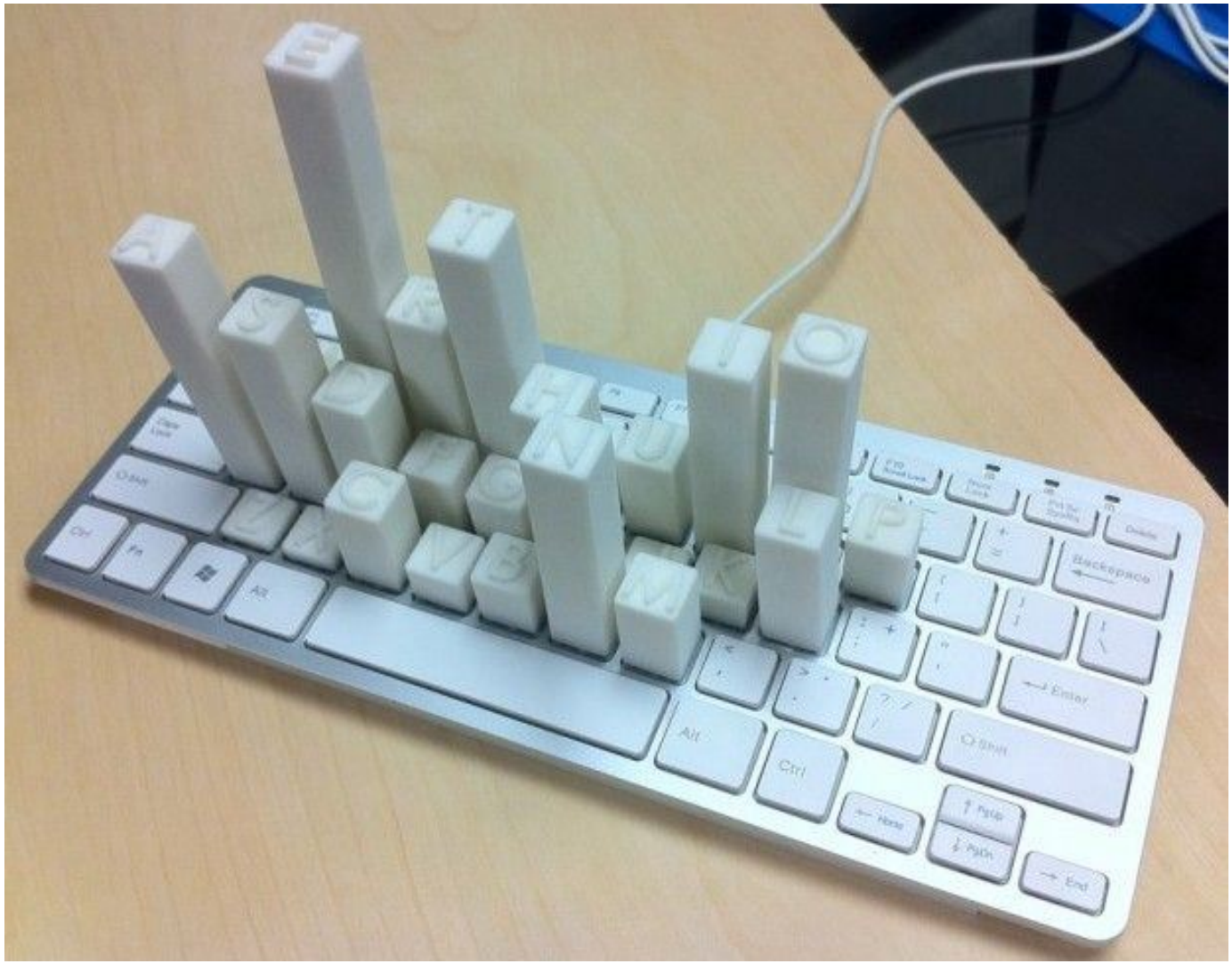
- 密码分析的一个基础是英语字母的统计特性。

英语字母的统计特征

- 密码分析的一个基础是英语字母的统计特性。



英语字母的统计特征(续)



英语里双字母组的出现频率

- 下面列出了英语里最常见的双字母组，据Cornell University Math Explorer's Project，该项目统计了不少于40,000 单词。

th 1.52%	en 0.55%	ng 0.18%
he 1.28%	ed 0.53%	of 0.16%
in 0.94%	to 0.52%	al 0.09%
er 0.94%	it 0.50%	de 0.09%
an 0.82%	ou 0.50%	se 0.08%
re 0.68%	ea 0.47%	le 0.08%
nd 0.63%	hi 0.46%	sa 0.06%
at 0.59%	is 0.46%	si 0.05%
on 0.57%	or 0.43%	ar 0.04%
nt 0.56%	ti 0.34%	ve 0.04%
ha 0.56%	as 0.33%	ra 0.04%
es 0.56%	te 0.27%	ld 0.02%
st 0.55%	et 0.19%	ur 0.02%

汉字的统计特征

- 汉字频度表统计资料来源于清华大学。统计使用字数6763个（国标字符集）。范文合计总字数86405823个。使用率最高的500个汉字如下。

汉字的统计特征

- 汉字频度表统计资料来源于清华大学。统计使用字数6763个（国标字符集）。范文合计总字数86405823个。使用率最高的500个汉字如下。

的一国在人了有中是年和大业不为发会工经上地市要个产这出行作生家以成到日民来我部对进多全建他公开们场展时理新方主企资实学报制政济用同于法高长现本月定化加动合品重关机分力自外者区能设后就等体下万元社过前面农也得与说之员而务利电文事可种总改三各好金第司其从平代当天水省提商十管内小技位目起海所立已通入量子问度北保心还科委都术使明着次将增基名向门应里美由规今题记点计去强两些表系办教正条最达特革收二期并程厂如道际及西口京华任调性导组东路活广意比投决交统党南安此领结营项情解议义山先车然价放世间因共院步物界集把持无但城相书村求治取原处府.....

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

仿射密码

密码体制7 (仿射密码)

设$\mathcal{P}=\mathcal{C}=\mathbb{Z}_{26}$。密钥空间$\mathcal{K}=\{(a,b)\in\mathbb{Z}_{26}\times\mathbb{Z}_{26}:\gcd(a,26)=1\}$。对于密钥$K=(a,b)\in\mathcal{K}$和任意的$x,y\in\mathbb{Z}_{26}$，定义加密规则为

$$e_K(x)=(ax+b)\bmod 26,$$

解密规则定义为

$$d_K(y)=a^{-1}(y-b)\bmod 26.$$

仿射密码的破译

1. 统计密文字母中出现的频数。
2. 根据英语文本的统计特性，猜测至少两对可能的明密文对：

$$e_K(x_1) = y_1, \quad e_K(x_2) = y_2.$$

3. 根据猜测的明密文对，列出二元一次方程组，

$$\begin{aligned} y_1 &= ax_1 + b \bmod 26 \\ y_2 &= ax_2 + b \bmod 26 \end{aligned}$$

   解出密钥$K = (a, b)$，其中x_1, x_2分别是密文y_1, y_2对应的明文字母。
4. 利用过程(3)解出的密钥解密所有密文，如果发现是意义的明文，那么这个密钥极有可能是正确的。
5. 若密钥正确则完成破译，否则重复(2)(3)(4)。

已知明文攻击下的仿射密码破译

- 在已知明文攻击下，不用进行密文字母频数统计工作，有条件直接获得两对正确的明密文对，从而使整个破译工作变得非常简单。

已知明文攻击下的仿射密码破译

- 在已知明文攻击下，不用进行密文字母频数统计工作，有条件直接获得两对正确的明密文对，从而使整个破译工作变得非常简单。

1. 根据已知的明密文对

$$e_K(x_1) = y_1, \quad e_K(x_2) = y_2.$$

列出二元一次方程组，

$$\begin{aligned} y_1 &= ax_1 + b \bmod 26 \\ y_2 &= ax_2 + b \bmod 26 \end{aligned}$$

2. 解出密钥$K = (a, b)$，完成破译。

代换密码的破译

- 代换密码的破译完全依赖于密文字母频数和英语文本的统计特性。同时，对英语语言的熟悉程度很大程度上影响破译工组的进度。

代换密码的破译

- 代换密码的破译完全依赖于密文字母频数和英语文本的统计特性。同时，对英语语言的熟悉程度很大程度上影响破译工组的进度。
- 在已知明文攻击下，有条件直接获得个别或一些明文字母对应的密文字母，从而直接掌握密钥的部分信息，大大简化破译工组。

代换密码的破译

- 代换密码的破译完全依赖于密文字母频数和英语文本的统计特性。同时，对英语语言的熟悉程度很大程度上影响破译工组的进度。
- 在已知明文攻击下，有条件直接获得个别或一些明文字母对应的密文字母，从而直接掌握密钥的部分信息，大大简化破译工组。
- 例如，获悉明文字母“the”对应的密文字母是“JNZ”，就意味着密文中字母“J”都可以用“t”来替换，字母“N”都可以用“h”来替换，字母“Z”都可以用“e”来替换。

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

维吉尼亚密码

密码体制8 (维吉尼亚密码)

设$\mathcal{P}=\mathcal{C}=\mathcal{K}=(\mathbb{Z}_{26})^m$。对于密钥$K=(k_1,k_2,\cdots,k_m)$和任意的

$$x=(x_1,x_2,\cdots,x_m)\in(\mathbb{Z}_{26})^m, y=(y_1,y_2,\cdots,y_m)\in(\mathbb{Z}_{26})^m,$$

定义加密规则为

$$e_K(x_1,x_2,\cdots,x_m)=(x_1+k_1,x_2+k_2,\cdots,x_m+k_m),$$

解密规则定义为

$$d_K(y_1,y_2,\cdots,y_m)=(y_1-k_1,y_2-k_2,\cdots,y_m-k_m),$$

并且以上运算都在$\mathbb{Z}_{26}$上进行。

维吉尼亚密码

密码体制8 (维吉尼亚密码)

设$\mathcal{P}=\mathcal{C}=\mathcal{K}=(\mathbb{Z}_{26})^m$。对于密钥$K=(k_1,k_2,\cdots,k_m)$和任意的

$$x=(x_1,x_2,\cdots,x_m)\in(\mathbb{Z}_{26})^m, y=(y_1,y_2,\cdots,y_m)\in(\mathbb{Z}_{26})^m,$$

定义加密规则为

$$e_K(x_1,x_2,\cdots,x_m)=(x_1+k_1,x_2+k_2,\cdots,x_m+k_m),$$

解密规则定义为

$$d_K(y_1,y_2,\cdots,y_m)=(y_1-k_1,y_2-k_2,\cdots,y_m-k_m),$$

并且以上运算都在$\mathbb{Z}_{26}$上进行。

- 破译维吉尼亚密码时，直接统计密文字母中出现的频数是否有效？

维吉尼亚密码的破译：确定密钥长度之Kasiski法

- Kasiski测试法可以用来确定维吉尼亚密密钥字的长度m。它基于这样一个事实，如果在密文中观察到两个相同的长度至少为3的密文段，那么这两个密文段很有可能是相同一个明文段的加密结果。

维吉尼亚密码的破译：确定密钥长度之Kasiski法

- Kasiski测试法可以用来确定维吉尼亚密密钥字的长度m。它基于这样一个事实，如果在密文中观察到两个相同的长度至少为3的密文段，那么这两个密文段很有可能是相同一个明文段的加密结果。

定义3 (Kasiski测试法)

搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$，即

$$m = \gcd(\delta_1, \delta_2, \cdots).$$

维吉尼亚密码的破译：确定密钥长度之Kasiski法

- Kasiski测试法可以用来确定维吉尼亚密密钥字的长度m。它基于这样一个事实，如果在密文中观察到两个相同的长度至少为3的密文段，那么这两个密文段很有可能是相同一个明文段的加密结果。

定义3 (Kasiski测试法)

搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$，即

$$m = \gcd(\delta_1, \delta_2, \cdots).$$

- 想一想为什么要“长度至少为3”？

维吉尼亚密码的破译：确定密钥长度之重合指数法(1)

定义4 (重合指数)

设$\mathbf{x} = x_1x_2\cdots x_n$是一个含有$n$个字符的字符串。$\mathbf{x}$的重合指数定义为$\mathbf{x}$中两个随机元素相同的概率，即从$n$个字符随机选取两个元素它们相同的概率，记为$I_c(\mathbf{x})$。

维吉尼亚密码的破译：确定密钥长度之重合指数法(1)

定义4 (重合指数)

设$\mathbf{x} = x_1x_2\cdots x_n$是一个含有$n$个字符的字符串。$\mathbf{x}$的重合指数定义为$\mathbf{x}$中两个随机元素相同的概率，即从$n$个字符随机选取两个元素它们相同的概率，记为$I_c(\mathbf{x})$。

- 设$f_0, f_1, \cdots, f_{25}$分别表示字母$A, B, C, \cdots, Z$在字符串$\mathbf{x}$中的出现次数。对于每一个i满足$0 \le i \le 25$，共有$\binom{f_i}{2} = c_{f_i}^2$种方法使得所选取的两个字母皆为第$i$个字母。字符串$\mathbf{x}$的重合指数$I_c(\mathbf{x})$可以由下面的公式给出：

$$I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}.$$

维吉尼亚密码的破译：确定密钥长度之重合指数法(2)

假设字符串$\mathbf{x}$是标准的英语文本串，那么每一个字母出现的概率应该符合英语字母的统计特征，即为固定的$p_0, p_1, \cdots, p_{25}$，这样的字符串$\mathbf{x}$ 的重合指数应该为固定的

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065.$$

维吉尼亚密码的破译：确定密钥长度之重合指数法(2)

假设字符串$\mathbf{x}$是标准的英语文本串，那么每一个字母出现的概率应该符合英语字母的统计特征，即为固定的$p_0, p_1, \cdots, p_{25}$，这样的字符串$\mathbf{x}$ 的重合指数应该为固定的

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065.$$

- 单表代换密码加密后不改变重合指数，即明文串和密文串的重合指数相同，即标准英语文本串的重合指数0.065。

维吉尼亚密码的破译：确定密钥长度之重合指数法(2)

假设字符串$\mathbf{x}$是标准的英语文本串，那么每一个字母出现的概率应该符合英语字母的统计特征，即为固定的$p_0, p_1, \cdots, p_{25}$，这样的字符串$\mathbf{x}$ 的重合指数应该为固定的

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065.$$

- 单表代换密码加密后不改变重合指数，即明文串和密文串的重合指数相同，即标准英语文本串的重合指数0.065。
- 多表代换密码，如维吉尼亚密码，加密后改变了重合指数，即明文串和密文串的重合指数不同。

维吉尼亚密码的破译：确定密钥长度之重合指数法(3)

设维吉尼亚密码加密的字符串为$\mathbf{y} = y_1 y_2 \cdots y_n$。可以把字符串$\mathbf{y}$分割成$m$个长度相等的子串，分别为

$$\begin{aligned}
\mathbf{y}_1 &= y_1 y_{m+1} y_{2m+1} \cdots \\
\mathbf{y}_2 &= y_2 y_{m+2} y_{2m+2} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_i &= y_i y_{m+i} y_{2m+i} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_m &= y_m y_{2m} y_{3m} \cdots
\end{aligned}$$

维吉尼亚密码的破译：确定密钥长度之重合指数法(3)

设维吉尼亚密码加密的字符串为$\mathbf{y} = y_1 y_2 \cdots y_n$。可以把字符串$\mathbf{y}$分割成$m$个长度相等的子串，分别为

$$\begin{aligned}
\mathbf{y}_1 &= y_1 y_{m+1} y_{2m+1} \cdots \\
\mathbf{y}_2 &= y_2 y_{m+2} y_{2m+2} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_i &= y_i y_{m+i} y_{2m+i} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_m &= y_m y_{2m} y_{3m} \cdots
\end{aligned}$$

- 如果m是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$分别由密钥分量$k_1, k_1, \cdots, k_m$以移位密码加密的方式获得。它们的重合指数

维吉尼亚密码的破译：确定密钥长度之重合指数法(3)

设维吉尼亚密码加密的字符串为$\mathbf{y} = y_1 y_2 \cdots y_n$。可以把字符串$\mathbf{y}$分割成$m$个长度相等的子串，分别为

$$\begin{aligned}
\mathbf{y}_1 &= y_1 y_{m+1} y_{2m+1} \cdots \\
\mathbf{y}_2 &= y_2 y_{m+2} y_{2m+2} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_i &= y_i y_{m+i} y_{2m+i} \cdots \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_m &= y_m y_{2m} y_{3m} \cdots
\end{aligned}$$

- 如果m是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$分别由密钥分量$k_1, k_1, \cdots, k_m$以移位密码加密的方式获得。它们的重合指数应该均为标准英语文本串的重合指数，即大约为0.065。

维吉尼亚密码的破译：确定密钥长度之重合指数法(4)

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$是通过不同密钥分量以移位密码加密的方式获得的。英语文本原有的重合指数特性被破坏了，其重合指数应该不等于明文英语文本串的重合指数。它们看起来应该更为随机。

维吉尼亚密码的破译：确定密钥长度之重合指数法(4)

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$是通过不同密钥分量以移位密码加密的方式获得的。英语文本原有的重合指数特性被破坏了，其重合指数应该不等于明文英语文本串的重合指数。它们看起来应该更为随机。
- 对于一个完全的随机串$\mathbf{x} = x_1 x_2 \cdots x_n \cdots$，其重合指数应为：

$$I_c(\mathbf{x})$$

维吉尼亚密码的破译：确定密钥长度之重合指数法(4)

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$是通过不同密钥分量以移位密码加密的方式获得的。英语文本原有的重合指数特性被破坏了，其重合指数应该不等于明文英语文本串的重合指数。它们看起来应该更为随机。
- 对于一个完全的随机串$\mathbf{x} = x_1 x_2 \cdots x_n \cdots$，其重合指数应为：

$$I_c(\mathbf{x}) \approx 26 \cdot \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0.038 < 0.065.$$

维吉尼亚密码的破译：确定密钥长度之重合指数法(4)

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$是通过不同密钥分量以移位密码加密的方式获得的。英语文本原有的重合指数特性被破坏了，其重合指数应该不等于明文英语文本串的重合指数。它们看起来应该更为随机。
- 对于一个完全的随机串$\mathbf{x} = x_1 x_2 \cdots x_n \cdots$，其重合指数应为：

$$I_c(\mathbf{x}) \approx 26 \cdot \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0.038 < 0.065.$$

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$的重合指数应该明显小于标准英语文本串的重合指数0.065。

维吉尼亚密码的破译：确定密钥长度之重合指数法(4)

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$是通过不同密钥分量以移位密码加密的方式获得的。英语文本原有的重合指数特性被破坏了，其重合指数应该不等于明文英语文本串的重合指数。它们看起来应该更为随机。
- 对于一个完全的随机串$\mathbf{x} = x_1 x_2 \cdots x_n \cdots$，其重合指数应为：

$$I_c(\mathbf{x}) \approx 26 \cdot \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0.038 < 0.065.$$

- 如果m不是密钥字的长度，那么子串$\mathbf{y}_1, \mathbf{y}_1, \cdots, \mathbf{y}_m$的重合指数应该明显小于标准英语文本串的重合指数0.065。
- 利用重合指数方法，我们可以在Kasiski测试法的基础上进一步确认维吉尼亚密码的密钥字长度m。

维吉尼亚密码的破译：确定密钥

现在来确定$K = (k_1, k_2, \cdots, k_m)$。

维吉尼亚密码的破译：确定密钥

现在来确定$K = (k_1, k_2, \cdots, k_m)$。

- 设$f_0, f_1, \cdots, f_{25}$分别表示字母$A, B, C, \cdots, Z$在字符串$\mathbf{y}_i$中的出现次数。$\mathbf{y}_i$的长度记为$n' = n/m$。

维吉尼亚密码的破译：确定密钥

现在来确定$K = (k_1, k_2, \cdots, k_m)$。

- 设$f_0, f_1, \cdots, f_{25}$分别表示字母$A, B, C, \cdots, Z$在字符串$\mathbf{y}_i$中的出现次数。$\mathbf{y}_i$的长度记为$n' = n/m$。
- 字符串$\mathbf{y}_i = y_i y_{m+i} y_{2m+i} y_{3m+i} \cdots$中的每个字母是对应的明文字母移动$k_i$个位置得到的，因此有

$$\frac{f_{k_i}}{n'} \approx p_0, \frac{f_{k_i+1}}{n'} \approx p_1, \cdots, \frac{f_{k_i+25}}{n'} \approx p_{25}.$$

维吉尼亚密码的破译：确定密钥

现在来确定$K = (k_1, k_2, \cdots, k_m)$。

- 设$f_0, f_1, \cdots, f_{25}$分别表示字母$A, B, C, \cdots, Z$在字符串$\mathbf{y}_i$中的出现次数。$\mathbf{y}_i$的长度记为$n' = n/m$。
- 字符串$\mathbf{y}_i = y_i y_{m+i} y_{2m+i} y_{3m+i} \cdots$中的每个字母是对应的明文字母移动$k_i$个位置得到的，因此有

$$\frac{f_{k_i}}{n'} \approx p_0, \frac{f_{k_i+1}}{n'} \approx p_1, \cdots, \frac{f_{k_i+25}}{n'} \approx p_{25}.$$

- 猜测$k_i = g$，如果是正确的，那么一定有

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{g+i}}{n'} \approx \sum_{i=0}^{25} p_i^2 \approx 0.065,$$

否则，M_g应不等于0.065。

维吉尼亚密码的破译：确定密钥(续)

- 进一步，如果猜测$k_i = g$错误，则M_g还应小于0.065，即

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{g+i}}{n'} < 0.065.$$

- 这一判断基于下面的一个不等式的性质。

维吉尼亚密码的破译：确定密钥(续)

- 进一步，如果猜测$k_i = g$错误，则M_g还应小于0.065，即

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{g+i}}{n'} < 0.065.$$

- 这一判断基于下面的一个不等式的性质。

定理1

假设$p_1, p_2, \cdots, p_n$和$q_1, q_2, \cdots, q_n$均为概率分布且$p_1 \geq p_2 \geq \cdots \geq p_n$。令$q'_1, q'_2, \cdots, q'_n$为$q_1, q_2, \cdots, q_n$的任意置换，则

$$\sum_{i=1}^{n} p_i q'_i$$

当$q'_1 \geq q'_2 \geq \cdots \geq q'_n$时取最大值。

维吉尼亚密码的破译流程

维吉尼亚密码的破译流程

1. (Kasiski测试法)搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$。

维吉尼亚密码的破译流程

1. (Kasiski测试法)搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$。
2. (重合指数法确定密钥字长度)把密文字符串$\mathbf{y}$分割成m个长度相等的子串

$$\mathbf{y}_i = y_i y_{m+i} y_{2m+i} \cdots,$$

其中$i = 1, 2, \cdots, m$。分别计算$\mathbf{y}_i$的重合指数。它们的重合指数应该均为标准英语文本串的重合指数，即大约为0.065，否则密钥字长度不等于m。

维吉尼亚密码的破译流程

1. (Kasiski测试法)搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$。
2. (重合指数法确定密钥字长度)把密文字符串$\mathbf{y}$分割成m个长度相等的子串

$$\mathbf{y}_i = y_i y_{m+i} y_{2m+i} \cdots,$$

其中$i = 1, 2, \cdots, m$。分别计算$\mathbf{y}_i$的重合指数。它们的重合指数应该均为标准英语文本串的重合指数，即大约为0.065，否则密钥字长度不等于m。
3. (重合指数法确定密钥分量)猜测第i个密钥分量为g，即$k_i = g$。如果是正确的，那么应该有$M_g = \sum_{i=0}^{25} \frac{p_i f_{g+i}}{n'} \approx 0.065$，否则，$M_g$应不等于（小于）0.065。

维吉尼亚密码的破译流程

1. (Kasiski测试法)搜索长度至少为3的相同的密文段，记录这些相同密文段到起始点之间的距离$\delta_1, \delta_2, \cdots$，猜测这些$\delta$的最大公因数为密钥字长度$m$。
2. (重合指数法确定密钥字长度)把密文字符串$\mathbf{y}$分割成m个长度相等的子串

$$\mathbf{y}_i = y_i y_{m+i} y_{2m+i} \cdots,$$

其中$i = 1, 2, \cdots, m$。分别计算$\mathbf{y}_i$的重合指数。它们的重合指数应该均为标准英语文本串的重合指数，即大约为0.065，否则密钥字长度不等于m。
3. (重合指数法确定密钥分量)猜测第i个密钥分量为g，即$k_i = g$。如果是正确的，那么应该有$M_g = \sum_{i=0}^{25} \frac{p_i f_{g+i}}{n'} \approx 0.065$，否则，$M_g$应不等于（小于）0.065。
4. 重复利用步骤(3)的方法获得其余的密钥分量。

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

计算安全

定义5 (计算安全)

如果使用最好的方法破译一个密码体制至少需要N次操作，并且N是一非常大的数字，在现实中根本无法完成这么多次的操作，则可以说这一个密码体制是计算安全的。

计算安全

定义5 (计算安全)

如果使用最好的方法破译一个密码体制至少需要N次操作，并且N是一非常大的数字，在现实中根本无法完成这么多次的操作，则可以说这一个密码体制是计算安全的。

- 实际中的计算安全性都是在一些特定的攻击方式的基础上来衡量的，用一种攻击方式不能破译某种密码体制，但并不能表明其他类型的攻击方式也不能破译该密码体制。

计算安全

定义5 (计算安全)

如果使用最好的方法破译一个密码体制至少需要N次操作，并且N是一非常大的数字，在现实中根本无法完成这么多次的操作，则可以说这一个密码体制是计算安全的。

- 实际中的计算安全性都是在一些特定的攻击方式的基础上来衡量的，用一种攻击方式不能破译某种密码体制，但并不能表明其他类型的攻击方式也不能破译该密码体制。
- 例如，用穷尽密钥搜索攻击去破译代换密码可能难以实现，但并不能表明其他类型的攻击方式也不能破译代换密码。实际上，结合语言的统计特性容易破译代换密码。

计算安全

定义5 (计算安全)

如果使用最好的方法破译一个密码体制至少需要N次操作，并且N是一非常大的数字，在现实中根本无法完成这么多次的操作，则可以说这一个密码体制是计算安全的。

- 实际中的计算安全性都是在一些特定的攻击方式的基础上来衡量的，用一种攻击方式不能破译某种密码体制，但并不能表明其他类型的攻击方式也不能破译该密码体制。
- 例如，用穷尽密钥搜索攻击去破译代换密码可能难以实现，但并不能表明其他类型的攻击方式也不能破译代换密码。实际上，结合语言的统计特性容易破译代换密码。
- 以后将要学到的各种分组密码体制和各种Hash函数算法大都是计算安全的。

可证明安全

定义6 (可证明安全)

假设能够破译一个密码体制，并且利用破译该密码体制的方法可以解决某一个被认为困难的经过深入研究的（计算）数学问题，则可以说这一个密码体制是可证明安全的。

可证明安全

定义6 (可证明安全)

假设能够破译一个密码体制，并且利用破译该密码体制的方法可以解决某一个被认为困难的经过深入研究的（计算）数学问题，则可以说这一个密码体制是可证明安全的。

- 可证明安全的密码体制没有直接给出破译某个密码体制的难度，而是用破译密码的难度与解决数学难题的难度做类比，然后说明破译密码的难度。

可证明安全

定义6 (可证明安全)

假设能够破译一个密码体制，并且利用破译该密码体制的方法可以解决某一个被认为困难的经过深入研究的（计算）数学问题，则可以说这一个密码体制是可证明安全的。

- 可证明安全的密码体制没有直接给出破译某个密码体制的难度，而是用破译密码的难度与解决数学难题的难度做类比，然后说明破译密码的难度。
- 理想的公钥密码体制因该是可证明安全的。

无条件安全

定义7 (无条件安全)

即使攻击者有无限的计算资源（计算速度无限快，存储容量无限大）也无法破译一个密码体制，则可以说这一个密码体制是无条件安全的。

无条件安全

定义7 (无条件安全)

即使攻击者有无限的计算资源（计算速度无限快，存储容量无限大）也无法破译一个密码体制，则可以说这一个密码体制是无条件安全的。

- 无条件安全是最理想的密码体制，但真正实现无条件安全却是非常困难的。

无条件安全

定义7 (无条件安全)

即使攻击者有无限的计算资源（计算速度无限快，存储容量无限大）也无法破译一个密码体制，则可以说这一个密码体制是无条件安全的。

- 无条件安全是最理想的密码体制，但真正实现无条件安全却是非常困难的。
- 无条件安全的加密体制通常被称为是完善保密(perfect secrecy)的。

无条件安全

定义7 (无条件安全)

即使攻击者有无限的计算资源（计算速度无限快，存储容量无限大）也无法破译一个密码体制，则可以说这一个密码体制是无条件安全的。

- 无条件安全是最理想的密码体制，但真正实现无条件安全却是非常困难的。
- 无条件安全的加密体制通常被称为是完善保密(perfect secrecy)的。
- 移位密码、代换密码、仿射密码和维吉尼亚密码等古典密码体制都可以达到无条件安全，即完善保密，但在现实中无法真正实现。

内容提要

1. 密码分析的基础
2. 仿射密码和代换密码的破译
3. 维吉尼亚密码的破译
4. 密码体制的安全性

作业

编程实现维吉尼亚密码的破译流程，并破译附件txt文件中的一段由维吉尼亚密码加密得到的密文。要求：给出完整的实现代码，以及运行结果和分析过程, 10月31前交电子版.

谢谢！

杜育松
东校园北学院楼三楼(国家保密学院)A304室
15918768869
duyuong@mail.sysu.edu.cn