物联网安全

第三章：密码学基础——古典密码学

冀晓宇
浙江大学

USSLAB

课程概要

- **密码学发展历史**
- **密码学基础知识**
- **古典密码学算法**
  - 凯撒加密/位移加密
  - 仿射密码
  - 维吉尼亚密码

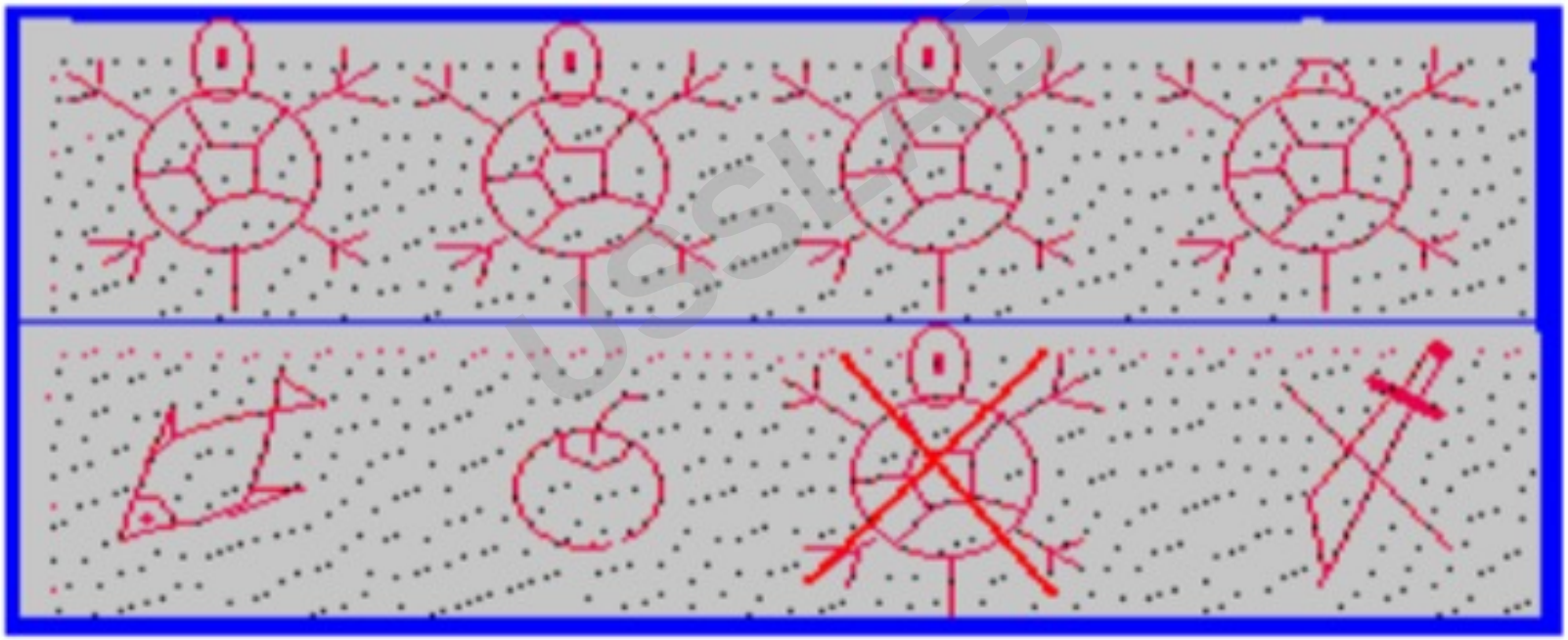
USSLAB

USSLAB

什么是密码学?

中国古代的密码艺术

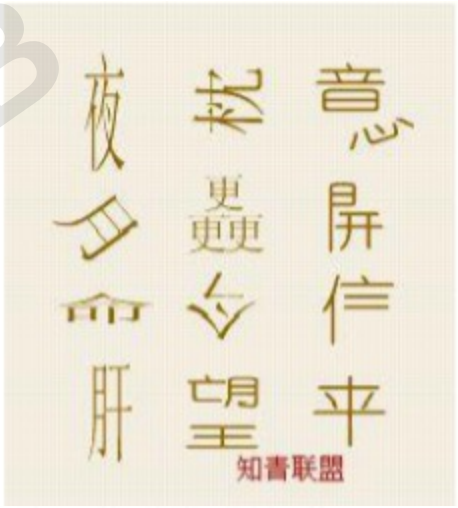
- 古代留守在家的妻子给外出工作的丈夫的书信



归，归，归！速归！如果（鱼果）不归，一刀两断

中国古代的密码艺术

- 会意诗

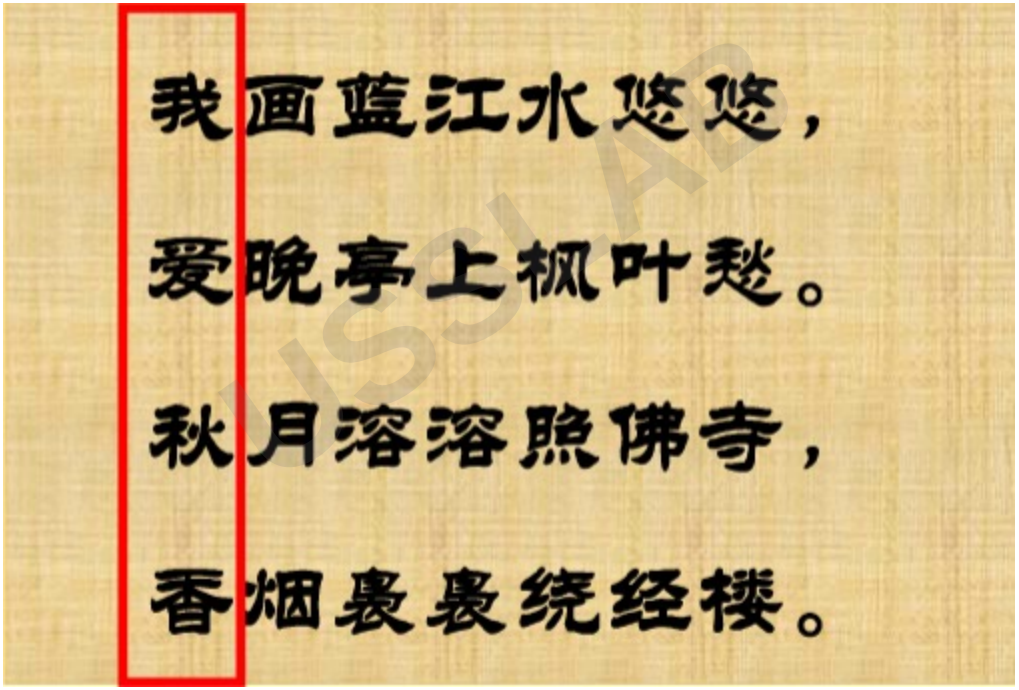


长夜横枕意心歪，
月斜三更门半开，
短命到今无口信，
肝肠望断无人来。

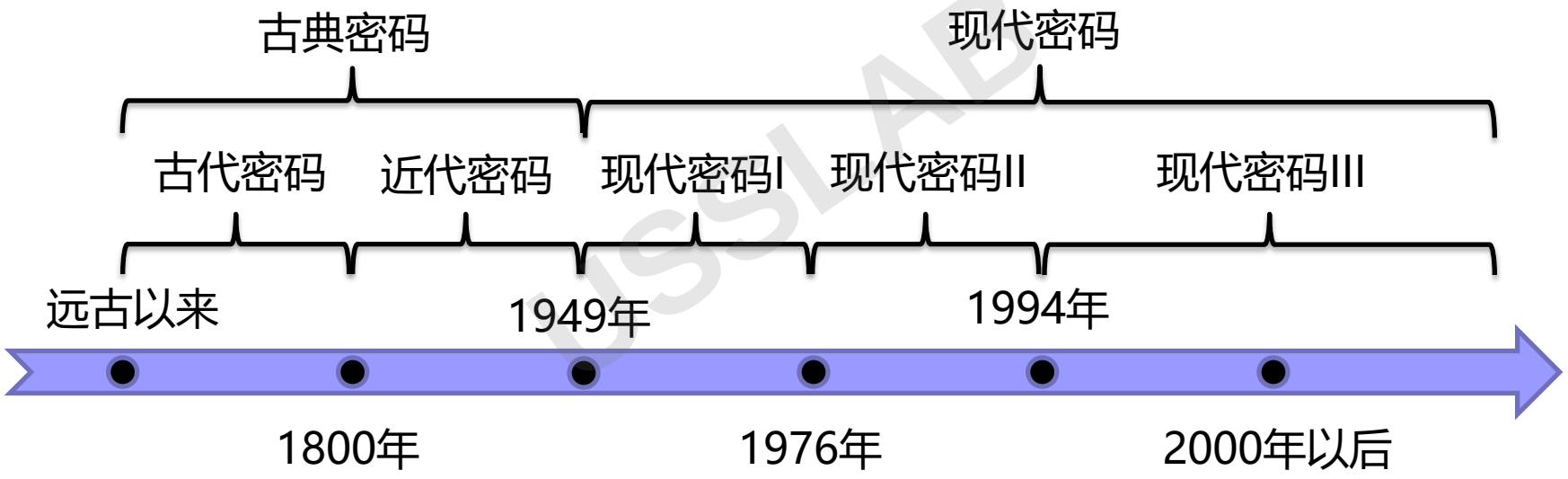
SLAB

中国古代的密码艺术

- 藏头诗



密码学的发展



USSLAB

古典密码阶段

- **时间：**
  1949之前
- **特点：**
  密码学还不是科学，而是艺术，但是出现了密码算法设计的基本哲学（**代替法&置换法**）
- **里程碑事件：**
  1883年科克霍夫第一次明确出密码编码原则

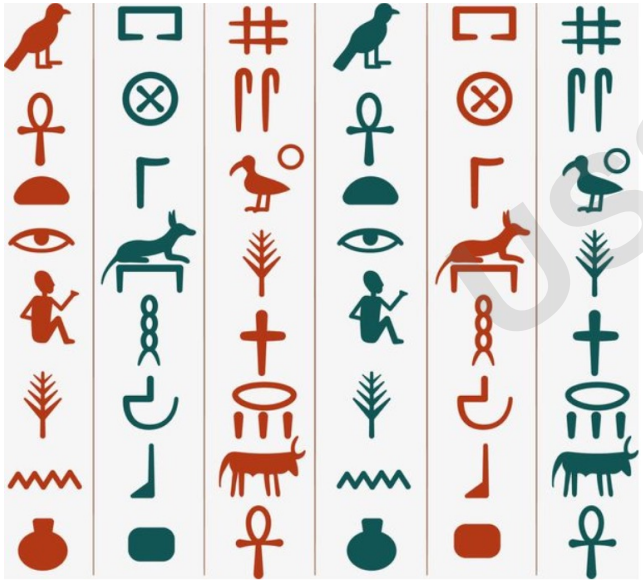
“加密算法应建立在算法公开不影响明文和密钥的安全的基础上，即安全性依赖于密钥的保密”

保密性：数据的保密**基于加密算法**的保密

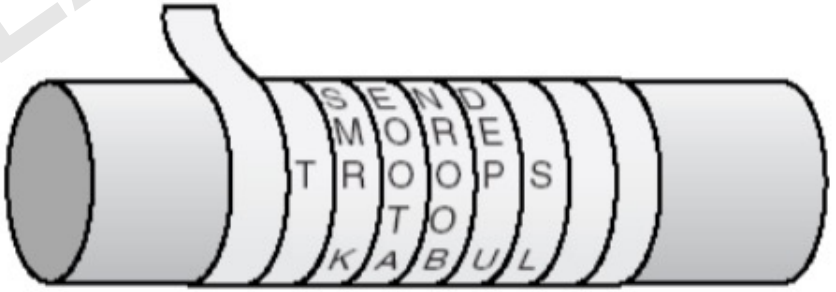
USSLAB

古典密码学哲学——代替和置换

- **代替（Substitution）：明文中的字符被替换成密文中另一个字符**
- **置换（Permutation）：明文字母不变，但顺序被打乱**



古埃及象形文字



古斯巴达人使用的天书

把长带子状的羊皮纸缠绕在圆木棍上，然后在上面写字；解下羊皮纸后，上面只有杂乱无章的字符，只有再次以同样的方式缠绕到同样粗细的棍子上

USSLAB

现代密码I阶段

- **时间：**
  1949-1976

- **特点：**
  - 密码学由艺术成为科学

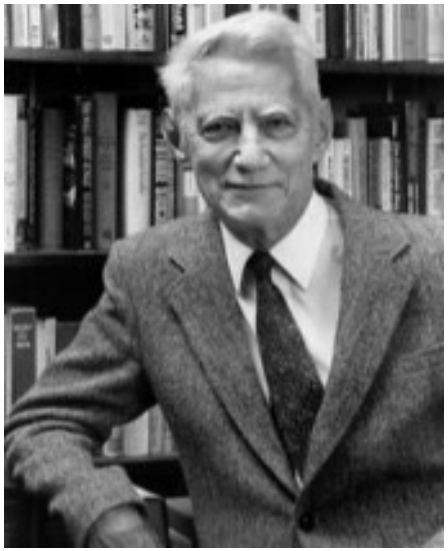
保密性：数据的**安全基于密钥**
而不是算法的保密

- **里程碑事件**
  - 1949年Shannon发表“The Communication Theory of Secret Systems”
  - 1967年，David Kahn专著《The Code Breakers》
  - 1971-1973年，IBM Waston实验室的Horst Feistel等人发表的几篇技术报告
  - 1974年，IBM提交了LUCIFER，后来成为DES

USSLAB

香浓的贡献——混淆和扩散

- 奠定了现代密码学的理论基础
- 定义了理论安全性，提出**扩散**和**混淆**原则（后续详细介绍）
  - **扩散**：将每一位明文尽可能地散布到多个输出密文中去，隐蔽明文数字的统计特性
  - **混淆**：使密文的统计特性与明文密钥之间的关系尽量复杂化



Shannon

USSLAB

现代密码学II阶段

- **时间：**

  1976-1994

- **特点：**
  - **公钥加密**出现

- **里程碑事件：**
  - 1976年Diffie&Hellman的“New Directions in Cryptography”提出了**公钥密码**的概念
  - 1977年Rivest, Shmir和Adleman提出了RSA公钥算法

Diffie&Hellman获2015图灵奖

USSLAB

现代密码III阶段

- **发展时间轴：**

1994年
Shor提出量子计算机模型下分解大整数和多项式时间算法

2000年
AES正式取代DES成为新的加密标准

2006年
第一届后量子密码学国际研讨会召开

2017年
NIST开始征集后量子密码标准

USSLAB

公钥密码未来发展——后量子公钥密码

- **后量子密码**：在传统计算机上运行，为了抵御量子计算对密码系统的威胁而产生的算法
- 目前的研究主要包括：
  - 基于编码的公钥密码
  - 基于格的公钥密码
  - 基于HASH的公钥密码
  - 多变量公钥密码

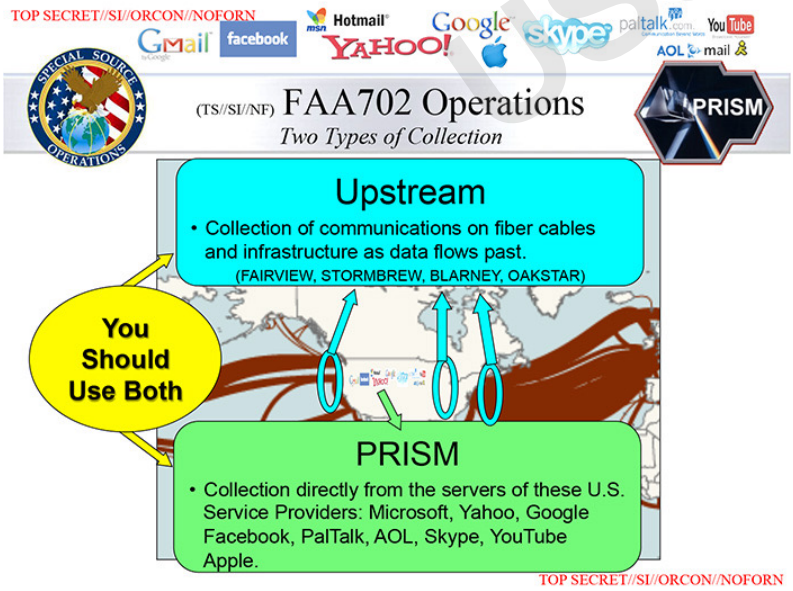
美国的“**上游**”计划：企图通过监听海底光缆截取流经海底光缆及通信基础设施的信息，以便量子计算机出现之后，进行开发。

USSLAB

趣闻：上游与棱镜

- **上游（UPSTREAM）**：海底光缆及基础设施监听计划。

- **棱镜（PRISM）**：是一项由美国国家安全局自2007年开始实施的绝密级网络监控计划。根据报导，泄露的文件中描述PRISM计划能够对即时通信和既存资料进行深度的监听。2013年，斯诺登将棱镜计划于香港透露给英国《卫报》。



斯诺登

USSTAO

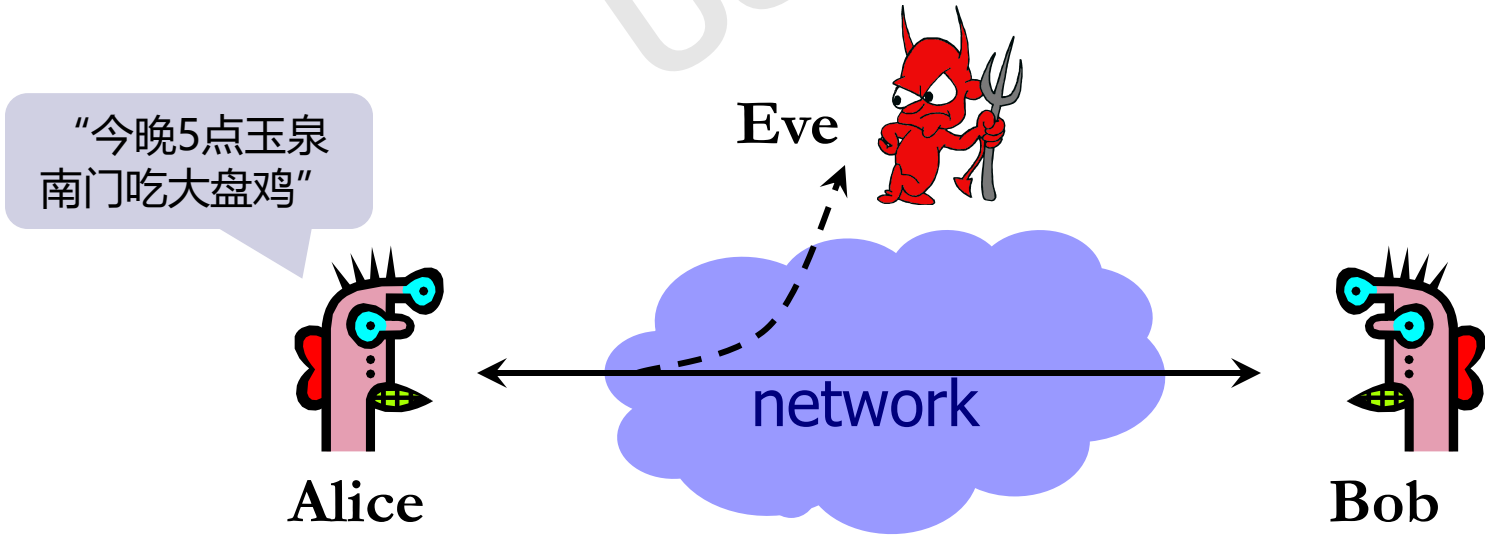
USSLAB

密码学基础知识

信息安全的目标

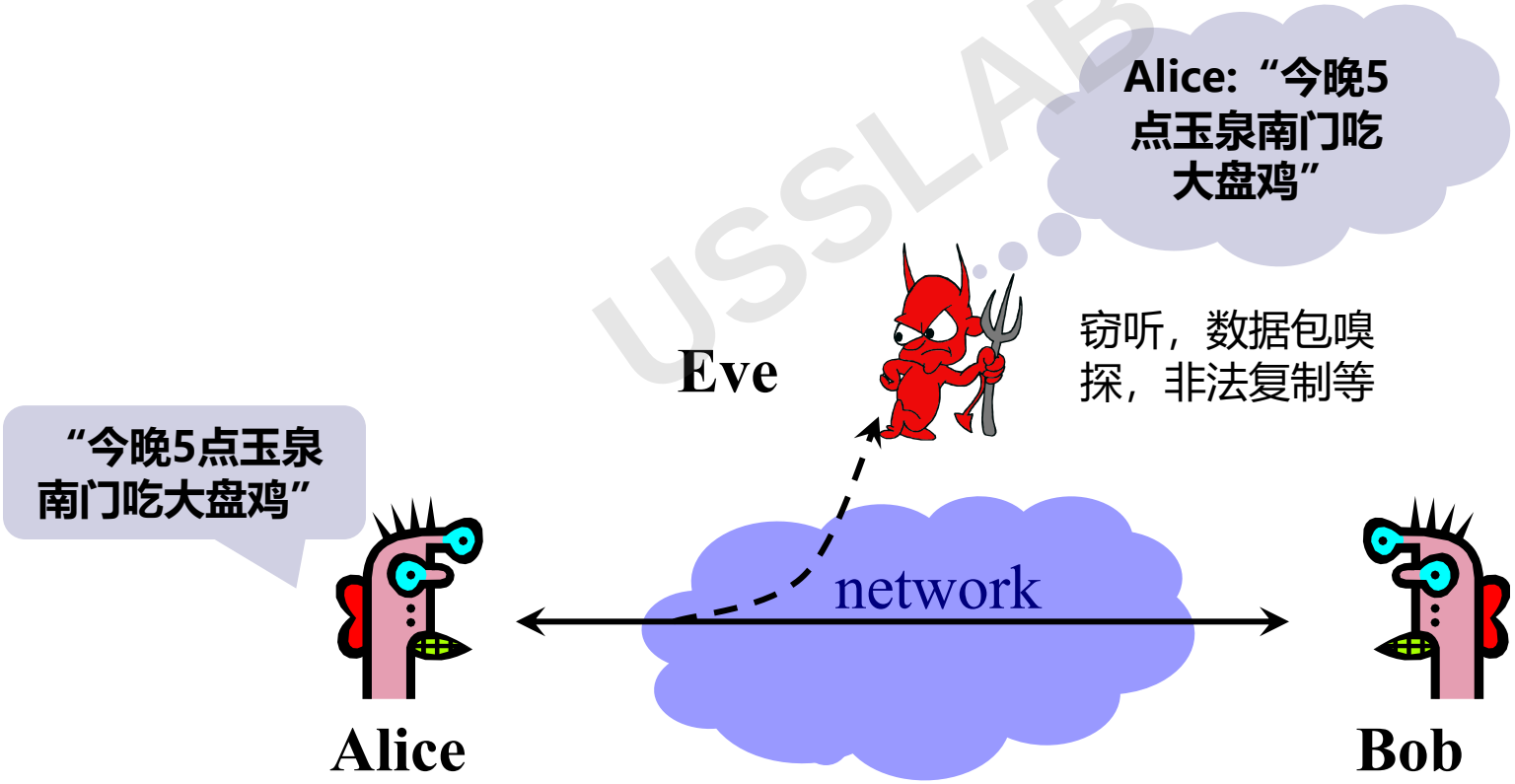
- 机密性（Confidentiality）：保证信息/数据不被泄露给未经授权的个体
- 完整性（Integrity）：保证信息/数据不会被未经授权的个体修改
- 可用性（Availability）：保证信息/数据能够被已授权的个体访问
- 真实性（Authenticity）：保证信息/数据确实来自其所声称的消息源
- 不可抵赖性（Non-repudiation）：保证信息/数据交互的所有参与者均不能否认曾经发送过的消息和数据

USSLAB



机密性（Confidentiality）

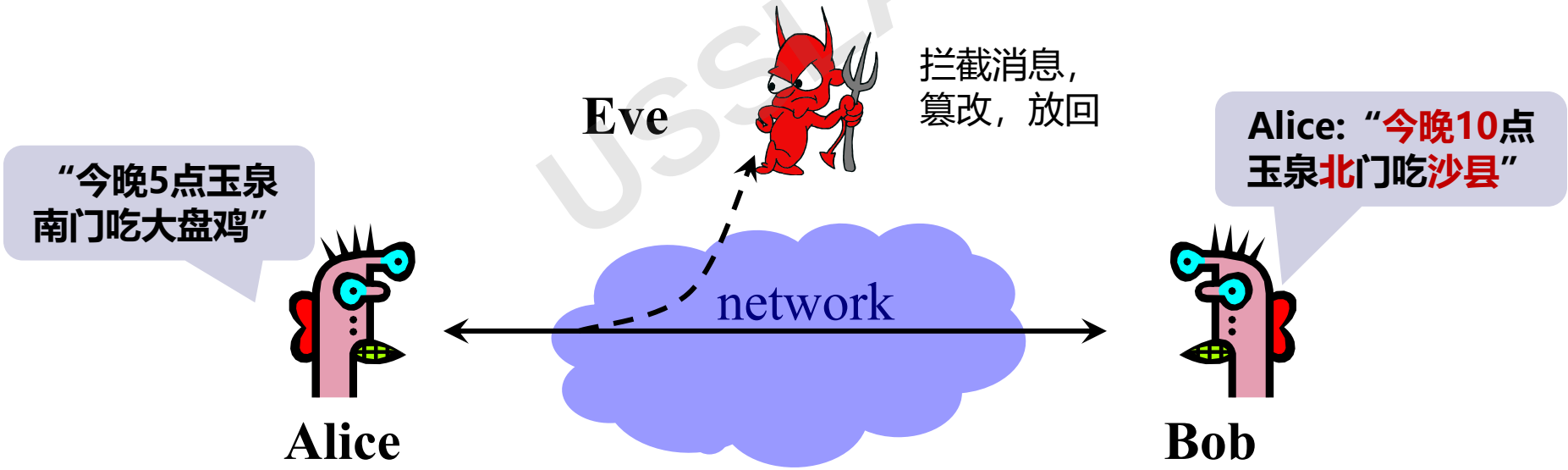
- 攻击者不能**解析**出来Alice给Bob的消息
- 防御方法：密码学加密解密算法



USSLAB

完整性（Integrity）

- Bob能够确定Alice的消息没有被**篡改**过
- 防御方法：Hash 函数、纠错码、消息认证码



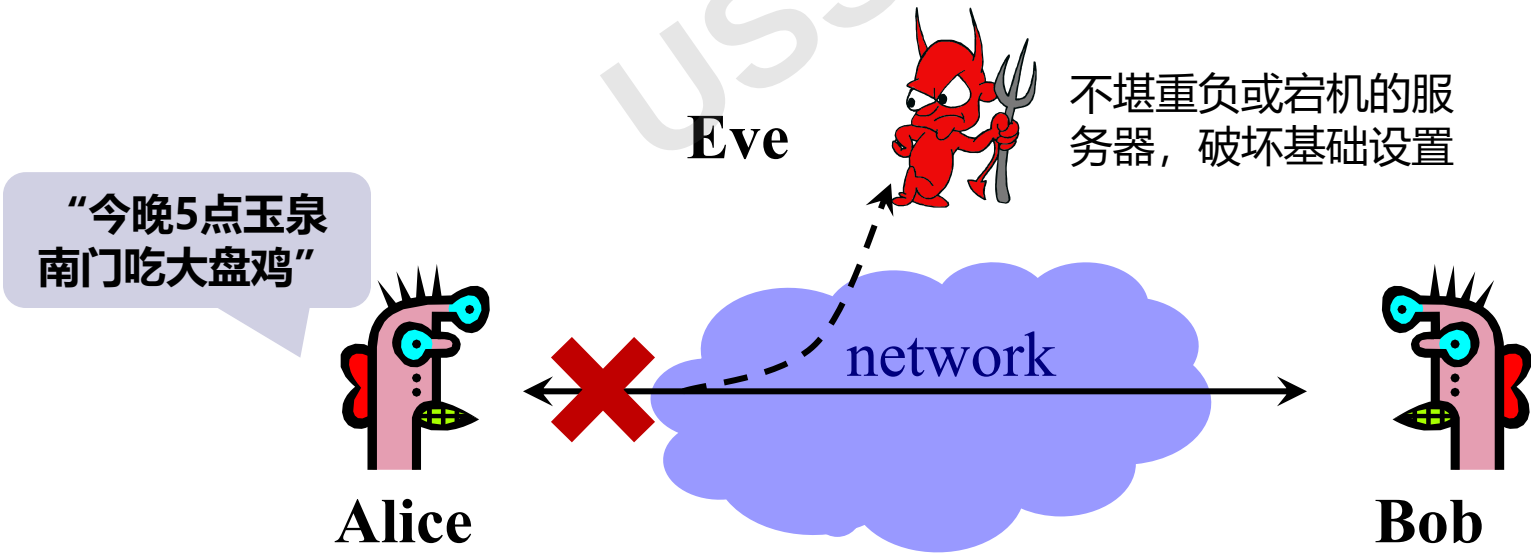
USSLAB

Q：机密性和完整性什么关系?

可用性（Availability）

- 可用性是用户**使用**需要的信息或资源的能力
- 破坏可用性的攻击：Denial of Service (Dos)、Distributed Dos (DDoS)

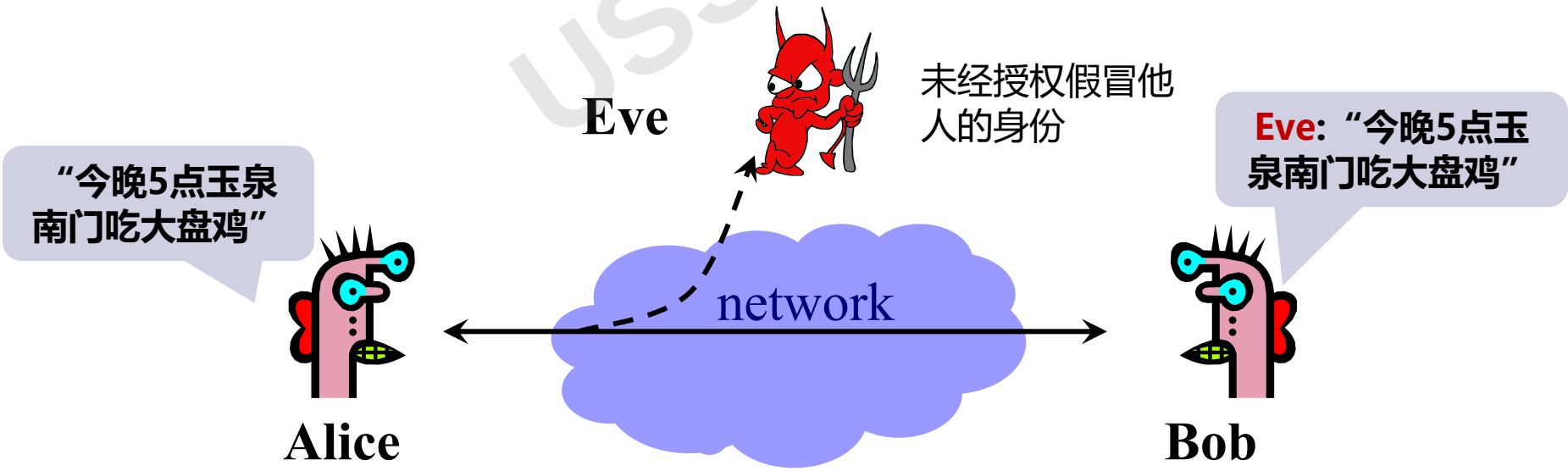
USSLAB



真实性（Authenticity）

- 能够识别/实体认证及确定**数据来源**
- 防御工具：密码学方案、数字签名、Hash函数、消息认证码、质询-响应协议

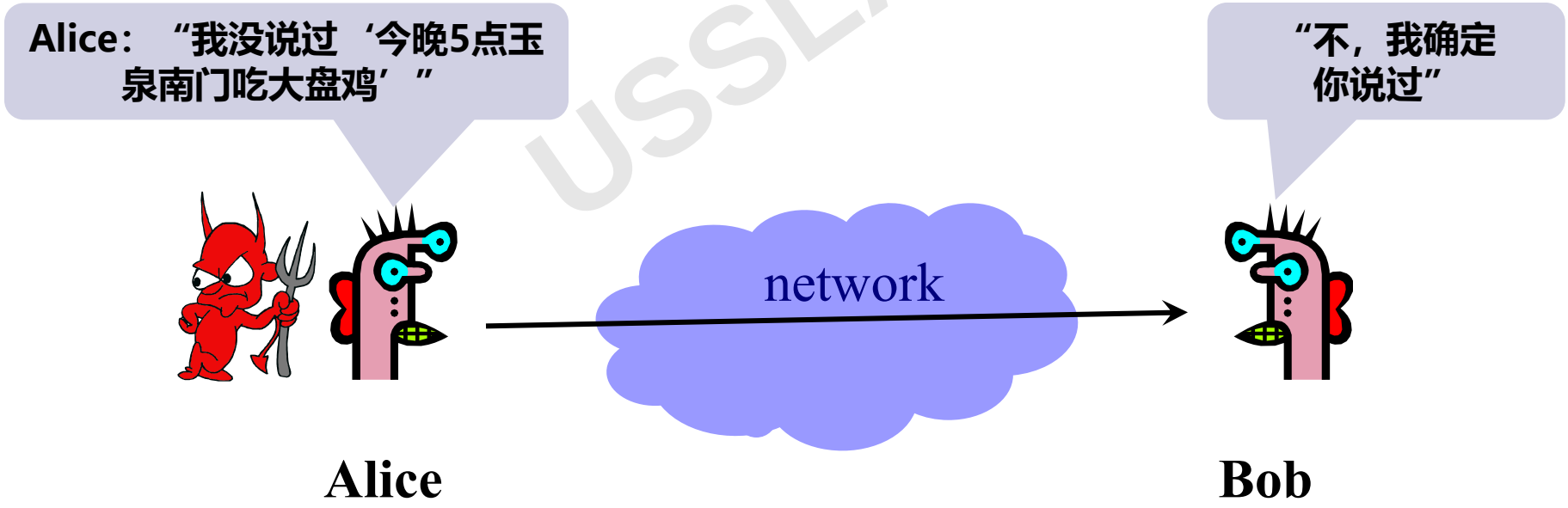
USSLAB



不可抵赖性（Non-repudiation）

- Alice不能声称自己没有发送消息
- 防御方法：数字签名

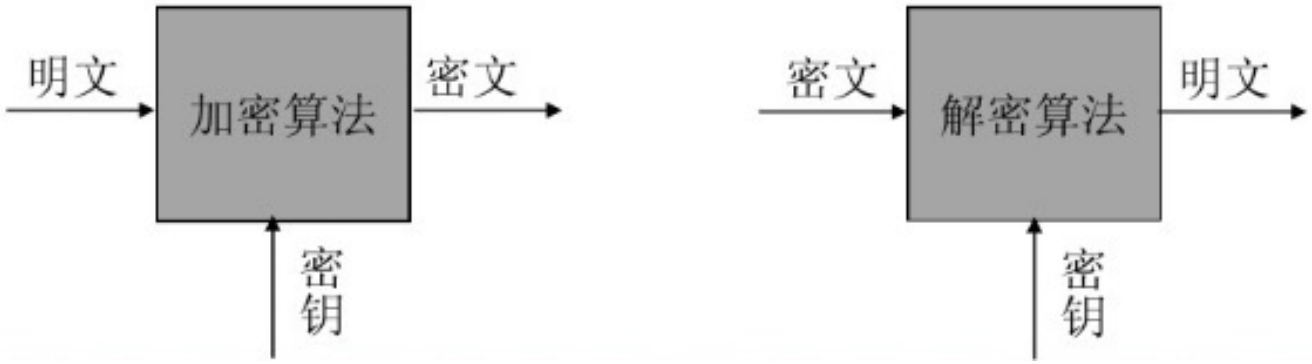
USSLAB



否认她发过消息

密码学基本概念

- 明文 Plaintext：需要秘密传送的信息
- 密文 Ciphertext：明文经过密码变换后的消息
- 加密 Encryption (E(m))：由明文到密文的变换
- 解密 Decryption (D(c))：从密文恢复出明文的过程
- 破译：非法接收者试图从密文分析出明文的过程
- 密钥：加密和解密时使用的一组秘密信息
- 加密/解密算法：对明/密文进行加/解密时的规则



USTAR

密码学安全的概念

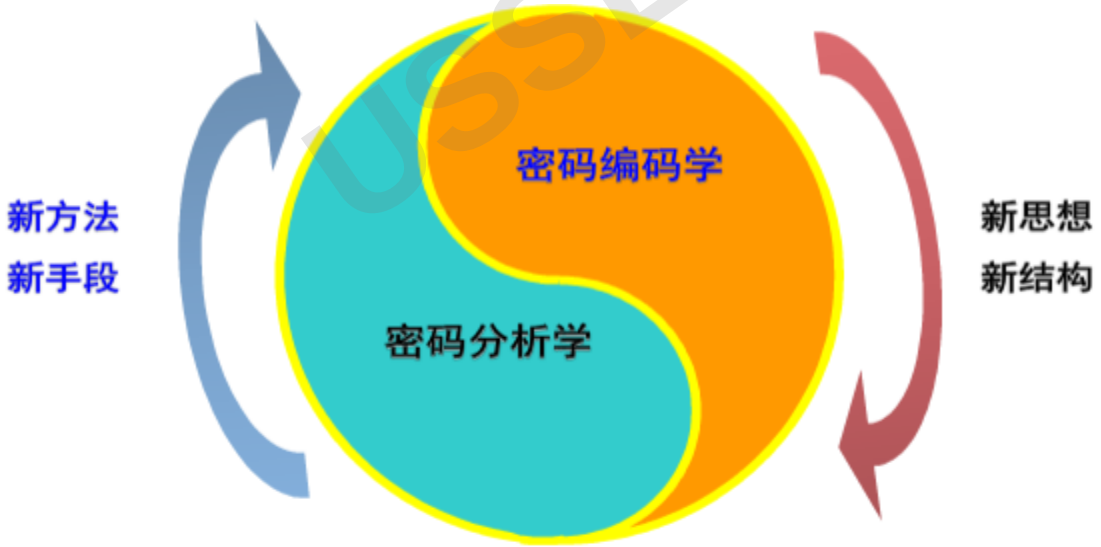
“如果把一封信缩在保险柜里，把保险柜藏起来，然后告诉你去看这封信，这并不是安全，而是隐藏；相反，如果把一封信锁在保险柜中，然后把**保险柜及其设计规范和许多同样的保险柜**给你，以便你和世界上最好的开保险柜的专家能够研究锁的装置，而你还是无法打开保险柜去读这封信，这才是安全”

——Bruce Schneier

USSLAB

密码学与密码分析学

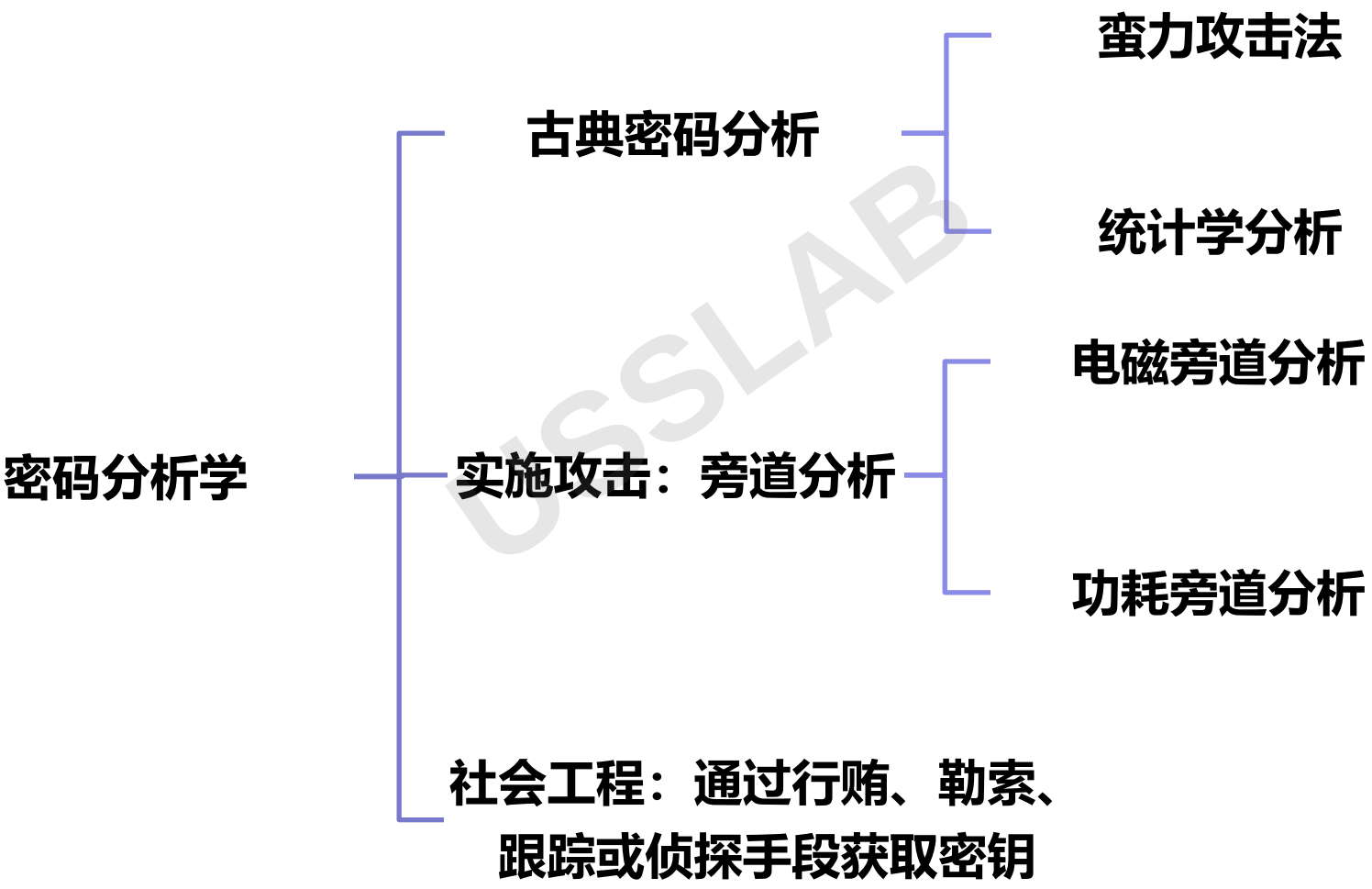
- **科克霍夫假设：**
  - 密码分析者知道双方使用的**密码系统**，包括明文的**统计特性、加密解密体质**等，唯一不知道的是密钥。
- **在设计一个密码系统时，目标是在科克霍夫的前提下实现安全**



两个对立统一的学科

USSLAB

密码分析学



USSLAB

案例1：侧信道攻击——加密密钥破解

[1] Genkin D, Pachmanov L, Pipman I, et al. Stealing keys from PCs using a radio: Cheap electromagnetic attacks on windowed exponentiation[C]//International workshop on cryptographic hardware and embedded systems. Springer, Berlin, Heidelberg, 2015: 207-228.
[2] Genkin D, Pachmanov L, Pipman I, et al. ECDH key-extraction via low-bandwidth electromagnetic attacks on PCs[C]//Cryptographers' Track at the RSA Conference. Springer, Cham, 2016: 219-235.

案例2：社会工程学

- 社会工程学：通过与他人的合法地交流，利用受害者**心理弱点**、本能反应、信任、好奇心、贪婪等心理使其心理受到影响，做出某些动作或者是透露一些机密信息的方式。

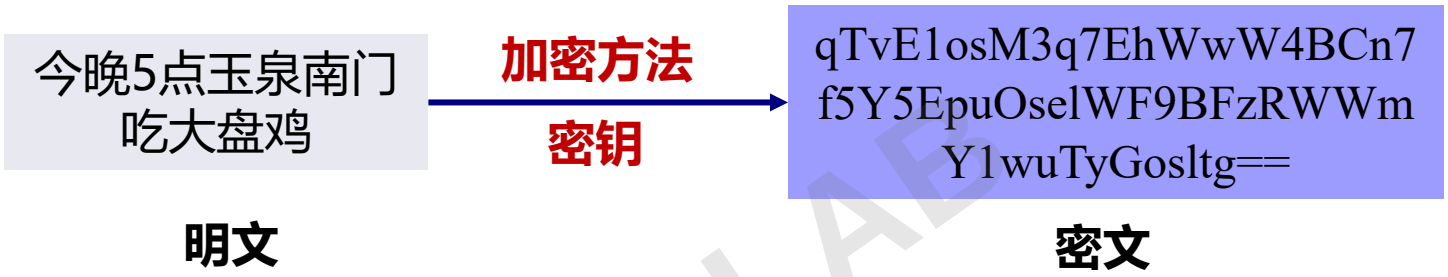
- 手段：通常以交谈、假托、调虎离山、在线聊天、下饵、等价交换、同情心、尾随等方式。

- 美国前头号黑客凯文·米特尼克被认为是社会工程学的大师和开山鼻祖，著有安全著作《欺骗的艺术》

USSLAB

密码学攻击分类

- **攻击的分类（以获取密钥和加解密算法为目的）**



USSLAB

难度由下到上逐渐增加

- **唯密文攻击（Cipher-only attack）**：攻击者**仅有**一个或多个密文，攻击需要统计分析（稍后详细介绍）；
- **已知明文攻击（Known-plaintext attack）**：攻击者有**一份**密文和对应的明文，进行算法和密钥推导；
- **选择明文攻击（Chosen-plaintext attack）**：攻击者有机会**使用加密机**，因此可以选择任何明文并产生对应的密文，成功概率更大；
- **选择密文攻击（Chosen-cipher attack）**：攻击者有机会使用解密机，因此可以选择一些密文并产生对应的明文。

密码算法安全性

- 1. 算法相关性

安全级别（Security Level）	算法
薄弱(Weak)	DES, MD5
传统(Legacy)	SHA-1
基准(Baseline)	3DES
标准(Standard)	AES-128,SHA-256
较高(High)	AES-192,SHA-384
超高(Ultra)	AES-256,SHA-512

USSLAB

密码算法安全性

- 2. 密钥长度

Bruce Schneier公开密钥长度建议值

年度	对于个人	对于公司	对于政府
1995	768	1280	1536
2000	1024	1280	1536
2005	1280	1536	2048
2010	1280	1536	2048
2015	1536	2048	2048

USSLAB

密钥长度越长，加密数据越不容易被非法解密

密码编码学：加密算法分类

USSLAB

3.2 Classic Cryptograph

古典密码学

USSLAB

凯撒大帝是如何指挥军队的?

下午3时，向敌军城堡进攻，兵力10万人

军令可能会被窃取

传令兵向军队传达指令

凯撒是如何让军令安全送达军队的?

USSLAB

3.2.1 移位加密（凯撒加密）

1. 由于英文字符有26个字母，可以建立英文字母和模26的剩余之间的对应关系：

$$A = 0 \quad B = 1 \quad C = 2 \quad \ldots \quad Y = 24 \quad Z = 25$$

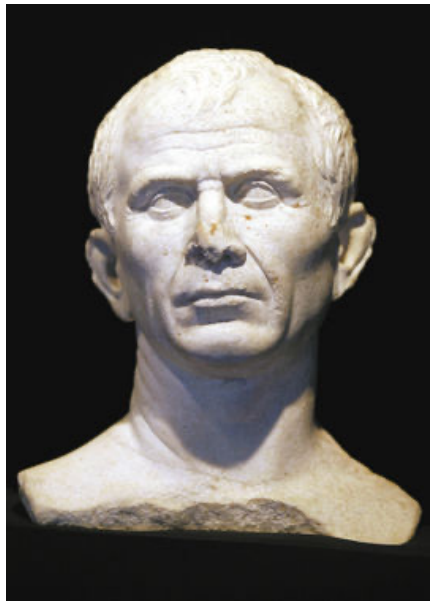
2. 加密过程:

$$y = x + k \pmod{26}$$

3. 解密过程

$$x = y - k \pmod{26}$$

4. 其中，**移位是加密方法，k是加密密钥**，k=3即为凯撒加密。



USSLAB

移位加密安全性

- 给定凯撒加密明文和密文，如下

明文字母	ABCDEFGHIJKLMNOPQRSTUVWXYZ
密文字母	DEFGHIJKLMNOPQRSTUVWXYZABC

  - 设明文为：LOVE， 则密文为：ORYH

- **想一想，如何攻击这种加密方式?**
  - 已知明文攻击：x,y都已知，那么

$$k = y - x \pmod{26}$$

  - 选择明文攻击：能够选择自己选择x。比如选择 x=‘a’=0 ，则为

$$y = k \pmod{26}$$

  - 选择密文攻击：自己选择密文y。例如选择 y=‘a’=0，那么

$$x = -k \pmod{26}$$

USSLAB

移位加密安全性（续）

- 如果只知道密文（唯密文）要如何攻击，例如，只知道密文文本“ORYH”？

- 暴力穷举：把26种可能都试一遍！

USSLAB

3.2.2 仿射加密

- **移位加密只有加减移位，易被暴力破解**
- **仿射加密引入乘法系数**

加密：给定密钥(α,β) $y = \alpha x + \beta \pmod{26}$

解密： $\alpha^{-1} \cdot (y-\beta) = x \bmod 26$

- 逆元：α^{-1}是α在模 26 下的**逆元**。例如，α=7, α逆元：

7*1 = 7 mod 26　　7*4 = 2 mod 26
7*2 = 14 mod 26　　7*5 = 9 mod 26
7*3 = 21 mod 26　　7*15=1 mod 26 …

得出 7*15=1 mod 26，**即15是7模26的逆元。**

Q：是不是每个数模26都存在逆元?

- **逆元存在条件**

α 必须与模数 26 **互质**，即 gcd (α,26)=1，逆元才存在。

USSLAB

仿射加密安全性

- **Q：仿射加密的密钥空间多大，即（α,β）组合多少种?**

- **限制1**：α如果相对模数n存在逆元，则需满足**gcd(α,n)=1**，也就是α与26必须互质，因此α的取值只有12种：

  {1，3，5，7，9，11，15，17，19，21，23，25}

- **限制2**：β取值范围是26。

- **结论**：对于仿射加密，所有可能只有12*26=312种！

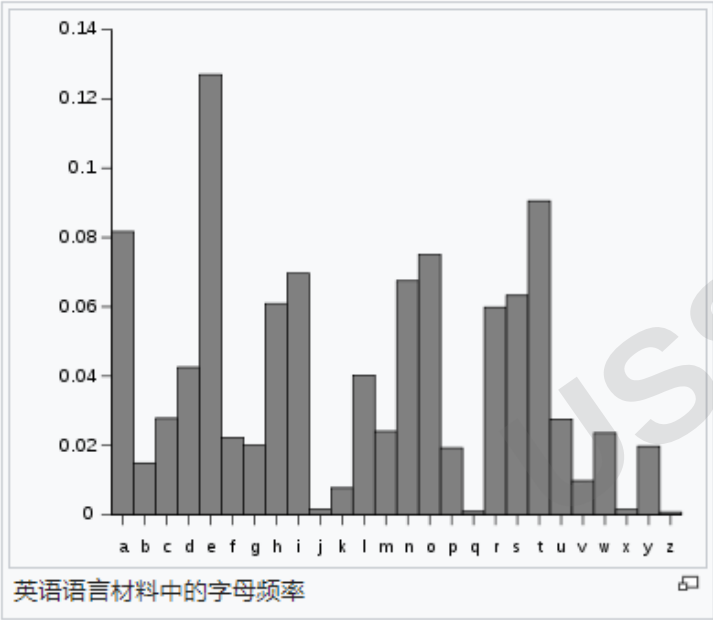
在当前计算机的计算能力下是非常不安全的

- **仿射加密的攻击：**
  - 暴力穷举：穷举出所有的312种可能
  - **词频统计攻击**

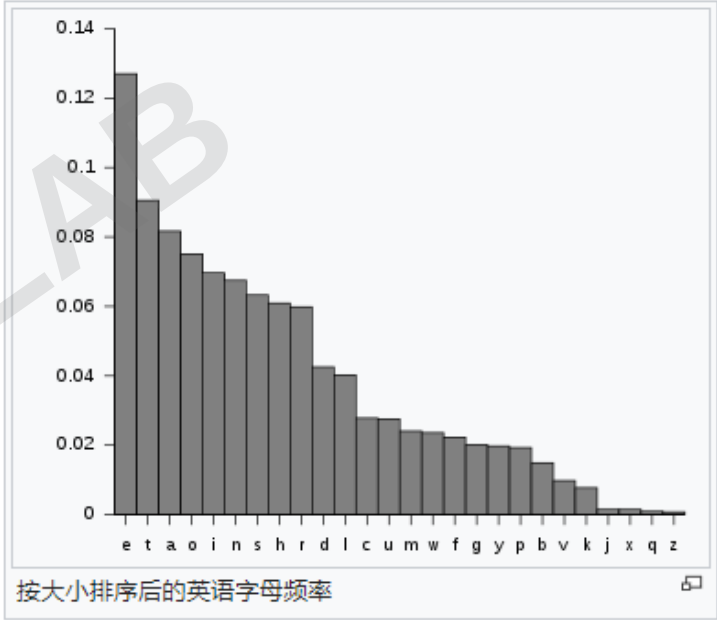
USSLAB

词频统计方法

- 英文字母概率统计分布



英语语言材料中的字母频率



按大小排序后的英语字母频率

USSLAB

- 词频统计方法
  - 统计密文中字母出现的频率
  - 与标准的语言字母出现的频率进行比对
  - 确定密钥的最可能值

仿射加密举例——大家猜猜看?

Pu yfo of oin hvy ufa hrpkpyb, jlar ph hopkk py oin hvy oinan, svo jnjpkk klvbi rfan zfyupgnyo zlkr; pu ovayng of ufvyg iph fjy hilgfj, lmmafmaplon nhzlmn, oin hvy jpkk sn oiafvbi oin inlao,jlar nlzi mklzn snipyg oin zfayna; pu ly fvohoanozing mlkr zlyyfo ulkk svoonaukx, oiny zknyzing jlcpyb larh, bpcny mfjna; pu P zly'o ilcn sapbio hrpkn, po jpkk ulzn of oin hvyhipyn, lyg hvyhipyn hrpkn ofbnoina, py uvkk skffr.

USSLAB

仿射加密破解

- **仿射加密是线性映射，所以明文和对应密文出现的频率是一致的。**

1. 统计密文中字母的频率

{'a': 18, 'c': 3, 'b': 7, 'g': 8, 'f': 19, 'i': 23, 'h': 17,
'k': 22, 'j': 10, 'm': 7, 'l': 21, 'o': 30, 'n': 37, 'p': 26,
's': 6, 'r': 10, 'u': 11, 'v': 13, 'y': 27, 'x': 1, 'z': 12}

2. 与标准字母频率比较，找到两个明文密文映射

分类	使用频率分类字母集	每个字母约占百分数
I	极高使用频率字母集:e	12%
II	次高使用频率字母集:t,a,o,i,n,s,h,r	6%~9%
III	中使用频率字母集:d,l	4%
IV	低使用频率字母集:c, u,m,w,f,g,y,p,b	1.5%~2.3%
V	次低使用频率字母集:v,k,j,x,q,z	1%

3. 确定映射关系

e→n; t→o

USSLAB

仿射加密

4. 按照映射关系求解密钥(α,β)破解密文

- 最高频和次高频的密文分别是n、o：e→n，t→o，得到方程组：

$$13 = 4\alpha + \beta \ （mod\ 26）$$

$$14 = 19\alpha + \beta \ （mod\ 26）$$

- 解出$\alpha = 7$，$\beta = 11$
- 求出α的逆元$\frac{1}{\alpha} = 15$， $X = (15\ Y - 11)\ mod\ 26$

USSLAB

If not to the sun for smiling warm is still in the sun there but we will laugh more confident calm if turned to found his own shadow appropriate escape the sun will be through the heart warm each place behind the corner if an outstretched palm cannot fall butterfly then clenched waving arms given power if I cant have bright smile it will face to the sunshine and sunshine smile together in full bloom

仿射加密

4. 按照映射关系求解密钥(α,β)破解密文

```
(7, 11, 'ifnottothesunforsmilingwarmisstillinthesuntherebutwe
(25, 17, 'cxtmddmdjekwtxmrkachctqigrackkdchhctdjekwtdjerezwdi
(3, 1, 'wpzknnknlecyzpkrcowdwzaumrowccnwddwznlecyznlerexynueu
(7, 11, 'ifnottothesunforsmilingwarmisstillinthesuntherebutwe
(9, 4, 'hwideedembjziwdojnhshirpvonhjjehsshiembjziembobqzepbp
(11, 14, 'tkilaalaqhxdikluxftctinjvuftxxatcctiaqhxdiaqhuhydaj
(15, 16, 'tcebmmbmwfpjecbsphtktezdrshtppmtkktemwfpjemwfsfojmc
(9, 2, 'ncojkkjkshpfocjuptnynoxvbutnppknyynokshpfokshuhwfkvhv
(3, 3, 'exhsvvsvtmkghxszkwelehicuzwekkvellehvtmkghvtmzmfgvcmc
(5, 21, 'eflcjjcjnosalfcbsuedelwiybuessjeddeljnosaljnobopajic
(7, 21, 'oltuzzuznkyatluxysorotmcgxsoyyzorrotznkyatznkxkhazck
```

如果不向太阳索取微笑,温暖仍在太阳那里,但我们会笑得更加自信从容;如果转过身去发现了自己的影子,适当的躲让,阳光便可穿越心灵,温暖每一处身后的角落;如果摊开的掌心不能点落蝴蝶,那就紧握成拳挥动臂膀,给予力量;如果我不能够微笑得灿烂,那就将脸投向灿烂的阳光,与阳光一起微笑,烂漫.

实际破解时，密钥(α,β)可能不止一组，需要根据密文的上下文场景和语义特征得到最可行的解法。

例子推导过程（课后习题）

- **1. 求解α的逆元：**

$$7*15=1 \bmod 26$$

$$1/\alpha=15$$

- **2. 根据 x = 1/α * (y-β) mod 26，已知1/ α = 15，可求得明文 x：**

USSLAB

密文	p	u	y	f	o	o	f
y	15	20	24	5	14	14	5
x=1/α (y-β)	60	135	195	-90	45	45	-90
x mod 26	8	5	24	5	14	14	5
明文	i	f	n	o	t	t	o

3.2.3 维吉尼亚密码

- **如何应对词频统计攻击?**
- **维吉尼亚密码**
  - 维吉尼亚密码是最早在1553年由吉奥万·巴蒂斯塔·贝拉索（Giovan Battista Bellaso）所著的书《吉奥万·巴蒂斯塔·贝拉索先生的密码》所创造。
  - 19世纪时被误认为是维吉尼亚发明的。

布莱斯·德·维吉尼亚

IISSLAB

维吉尼亚密码

- **加密方式**：
  - 列出明文并按照密钥长度分组
  - 用密钥对**每个组内**字母进行移位加密
  - 加密公式：C = (P+K) mod 26
- **特点**：
  - 维吉尼亚密码实际上移位密码的一种扩展
  - 能够**消除字母的频率特征**

Q:想一想为什么?

USSLAB

明文	H	e	r	e	i	s	h	o	w	i	t	i
密钥(长度6)	21	4	2	19	14	17	21	4	2	19	14	17
密文	C	I	T	X	W	J	C	S	Y	B	H	Z

维吉尼亚密码的安全性

- 破解维吉尼亚密码
  - 1. 找到密钥长度
  - 2. 找出密钥
- **找出密钥长度后破解密钥就很容易（Why?）：**
  - 把密文按照密钥长度L，选出密文中的第1个、第L+1、第2L+1个……字母进行词频统计。
  - 使用重合指数进行推测（后续详细讲述）
- 找到密钥长度的方法：
  - Kasiski实验
  - Friedman测试

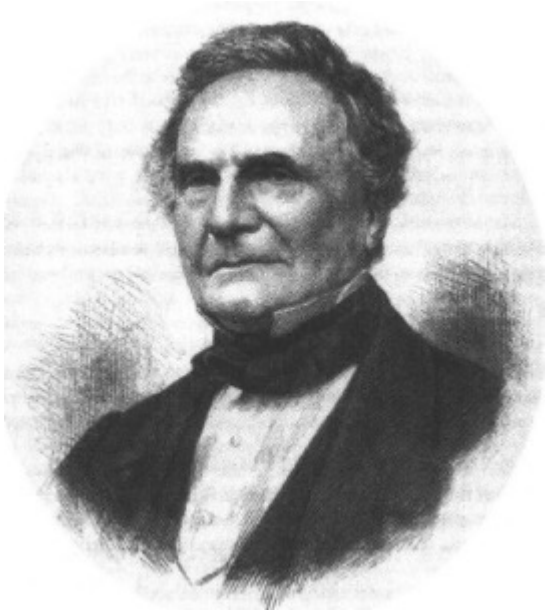
USSLAB

Kasiski-卡西斯基实验

密钥：FOREST FO RE STFO REST FO RES TFOR
明文：better to do well than to say well
密文：GSKXWK YC US OXQZ KLSG YC JEQ PJZC

12

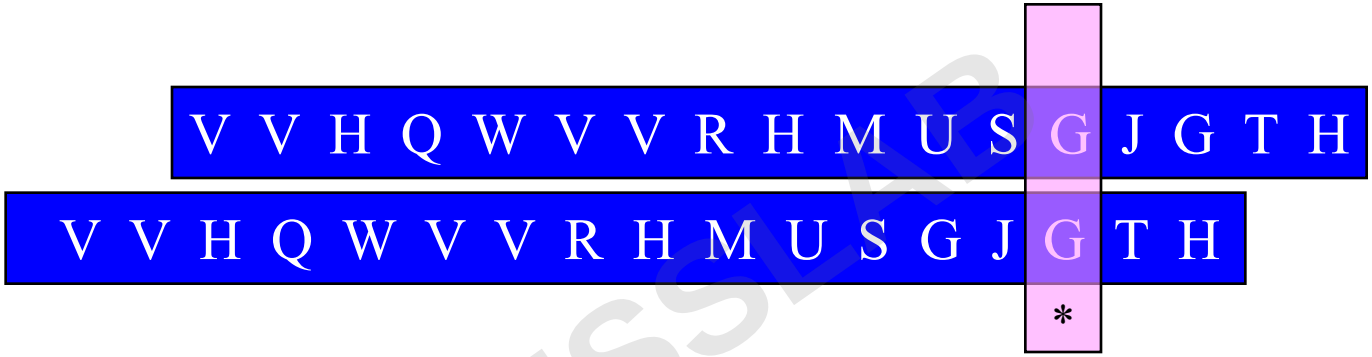
USSLAB



- 利用英文单词的规律，统计**重复间隔**。
- 如“YC”：间隔12。则公约数：1，2，3，4，6，12都可能是密钥长度。

找出密钥长度

1. 在两条纸上写下密文。将他们上下排好，两个纸条错开一定的距离。



2. 在两个相同字母的位置标上*

3. 改变两个纸条的错开位置，记录相同字母的个数

4. 出现最多相同字母的距离最可能是密钥长度

USSLAB

Friedman测试：基于重合指数计算

- IC, index of coincidence **重合指数**：同一份文本中随机选取两个字符相同的概率
- **IC计算方法**
  - 分子：计算**相同字母成对出现**的组合数
  - 分母：计算**所有可能的字母对**总数

语言	IC值
英语	0.0667
德语	0.0762
法语	0.0778
随机文本	0.0385

$$IC = \frac{\text{相同字母对数量}}{\text{总字母对数量}} = \frac{\sum C(n_i,2)}{C(N,2)} = \frac{\sum_{i=A}^{Z} n_i(n_i-1)}{N(N-1)}$$

符号	含义	示例（以英文单词"APPLE"为例）
n_i	第i个字母在文本中出现的次数	A出现1次，P出现2次，L、E各1次
N	文本总长度（所有字母数量）	N=5 ("A","P","P","L","E")
∑	对所有26个字母（A-Z）的统计值求和	计算所有字母的n_i（n_i-1）之和

Friedman测试：基于重合指数计算

- IC, index of coincidence **重合指数**：同一份文本中随机选取两个字符相同的概率
- **IC计算方法**
  - 分子：计算**相同字母成对出现**的组合数
  - 分母：计算**所有可能的字母对**总数

语言	IC值
英语	0.0667
德语	0.0762
法语	0.0778
随机文本	0.0385

$$IC = \frac{\text{相同字母对数量}}{\text{总字母对数量}} = \frac{\sum C(n_i,2)}{C(N,2)} = \frac{\sum_{i=A}^{Z} n_i(n_i-1)}{N(N-1)}$$

只有密钥长度L正确情况下，
密文IC值才会等于标准英文IC值（明文IC值）

示例：利用IC破解维吉尼亚密码

- **步骤1：推测密钥长度**

以密文VPXZGIAXIVWPUBTTMJPWIZITWZT为例

假设密钥长度	分组方式	平均IC值
3	第1/4/7...字母分组	0.042
4	第1/5/9...字母分组	**0.068**
5	第1/6/11...字母分组	0.039

结论：当分组长度=4时，IC值接近英语标准值0.066，推测真实密钥长度为4

- **步骤2：按密钥长度分组**

组1(第1,5, 9 ...位): V, G, I, P, M, I, T
组2(第2,6,10...位): P, I, V, U, J, Z, Z
组3(第3,7,11...位): X, A, W, B, W, T, W
组4(第4,8,12...位): Z, X, P, T, I, T, T

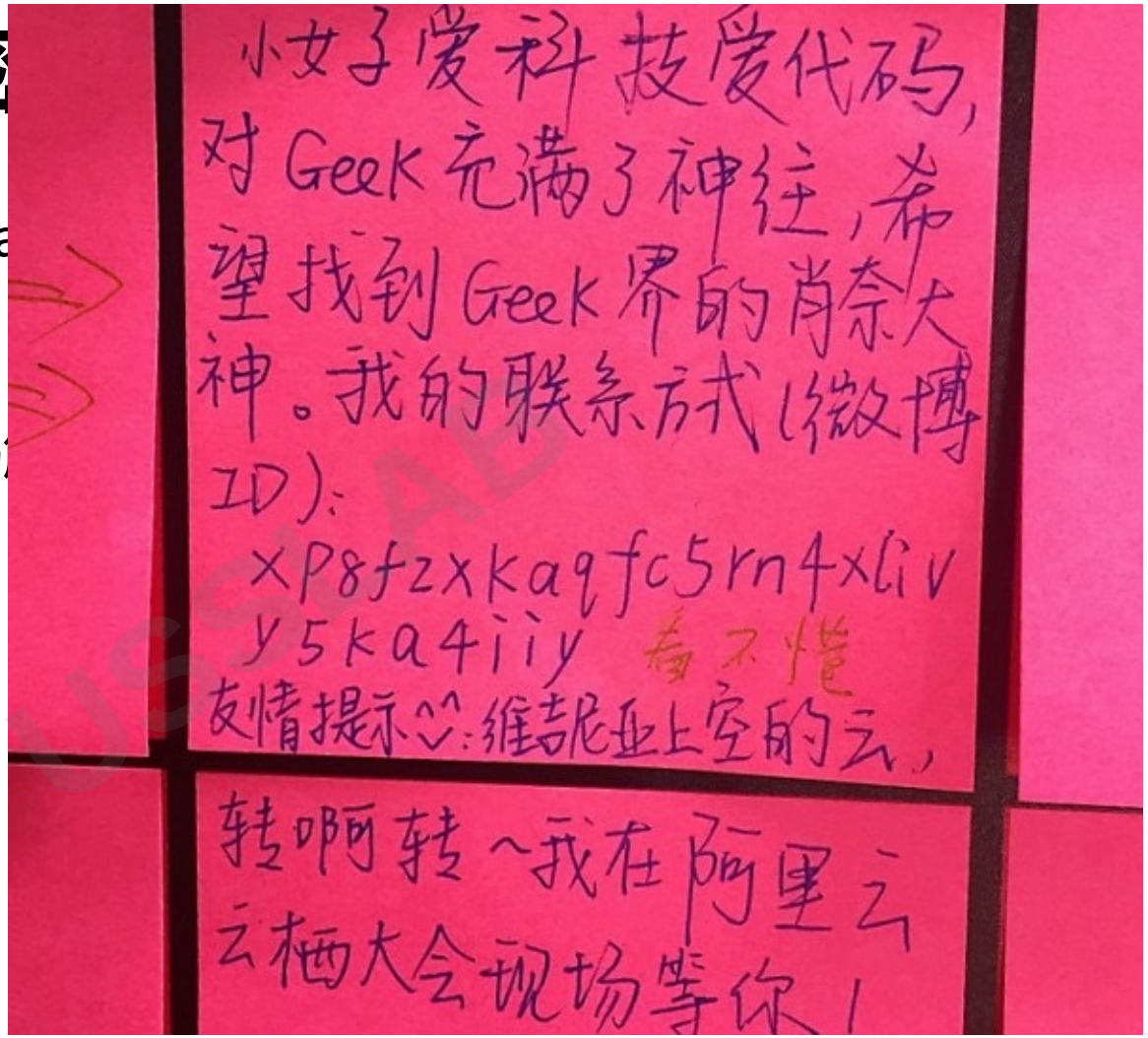
将密文拆分为4组（密钥长度=4）

USSLAB

维吉尼亚加密

- Xe8xbdxafxe5xa4xa
- 密钥：aliyun
- 方法：维吉尼亚密码
  +UTF-8中文解码

明文：软太太(微博)



演示程序☺

本章总结

- 密码学中信息安全目标
- 密码学体制分类
- 古典密码学思想
  - 凯撒加密、仿射加密、维吉尼亚加密的机制
- 古典密码学的安全性及攻击
  - 暴力破解、词频统计攻击等
- 古典密码学如何对现代密码学产生影响

USSLAB

本章结束

USSLAB