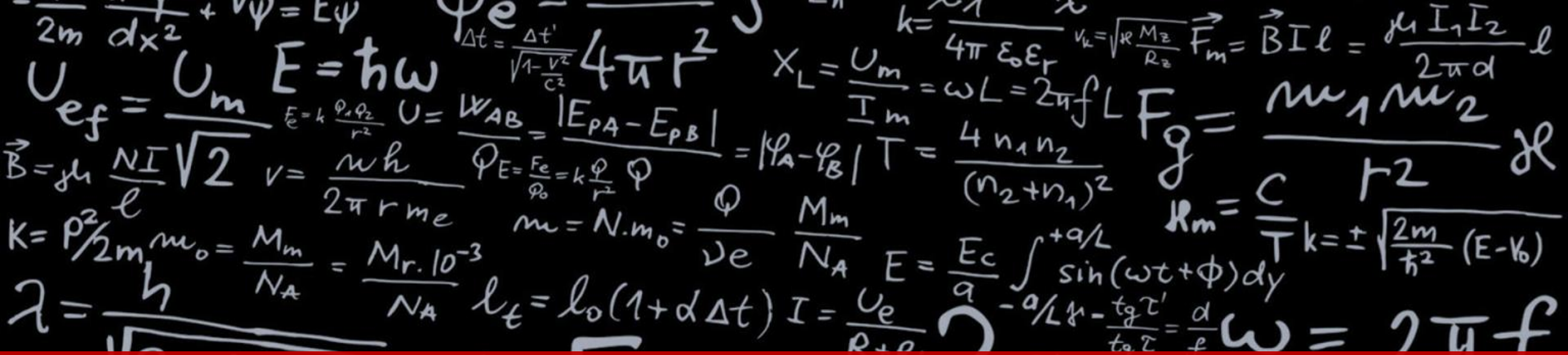


密码算法分析·引言

宋凌

邮箱：songlingcs@163.com

办公室：南海楼114

生活中的密码学：举例？

密码应用无处不在

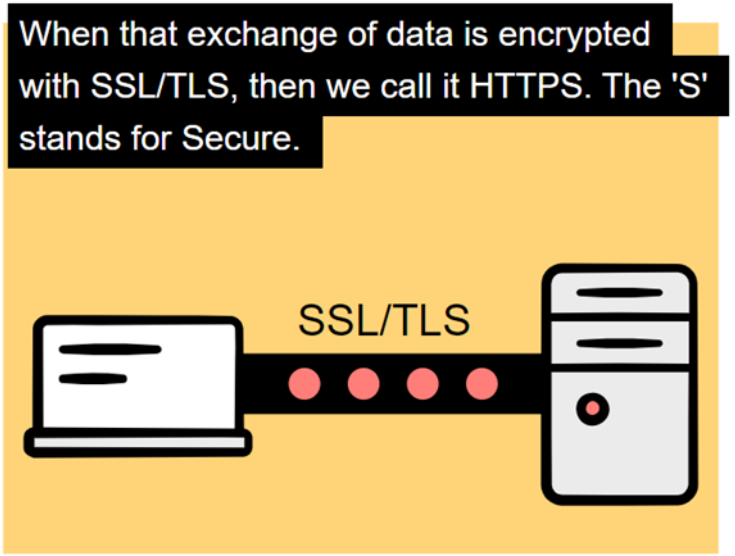
- 安全通信
  - web通信：https
  - 无线通信： 802.11i WPA2， GSM， Bluetooth
- 磁盘文件加密： EFS， TrueCrypt
- 数字内容保护(如 DVD, 蓝光光盘)： CSS，AACS
- 口令保护，身份认证，完整性校验，……

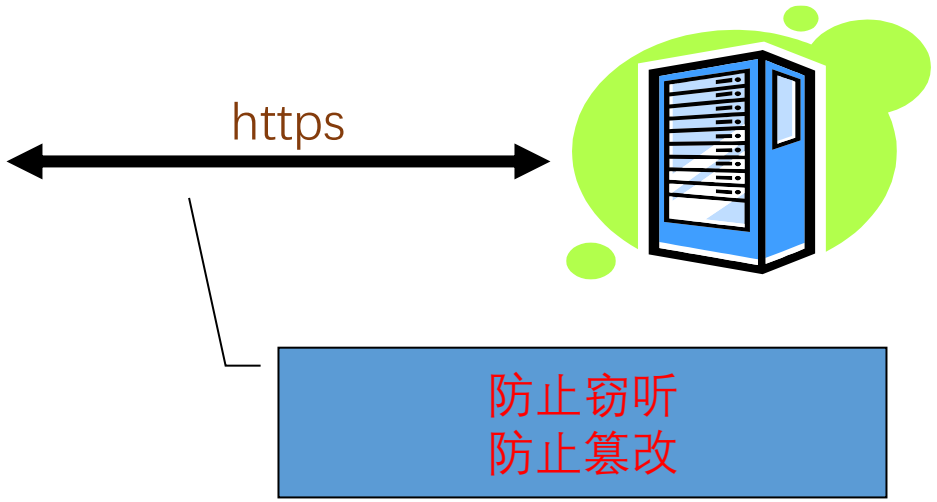
安全通信

- web通信： https
- 无线通信： 802.11i WPA2， GSM， Bluetooth

安全通信

- web通信： https
- 无线通信： 802.11i WPA2， GSM， Bluetooth

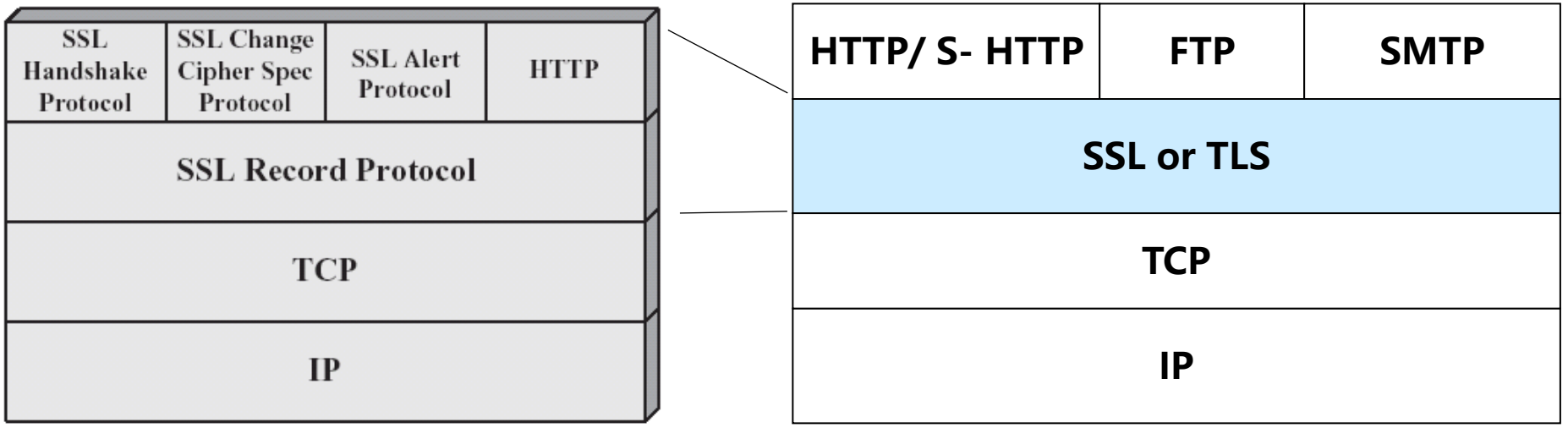




Secure Sockets Layer / TLS

两个主要部分

1. SSL 上层协议：**握手协议**
   - 使用公钥密码学协商一个秘密密钥
2. SSL 底层：**记录协议**
   - 用协商好的密钥，采用对称密码算法传输数据，确保机密性和完整性



磁盘文件加密

- EFS，Windows支持的一种文件加密属性
- TrueCrypt：一款免费开源的虚拟磁盘文件加密

身份认证：基于证书


现在有新的更新
单击以使用 Windows Update 安装这些更新。
下载XX安装包的请求
XX安装包，MS出品的证书

数字证书：把一个身份和一个公钥绑定

- 证书使用户相信，安装包的确来自微软。

数字证书

1. 版本号：XXXXXXXXX
2. 序列号：XXXXXXXXX
3. 签名算法标识
4. 颁发者
5. 有效期：XXXX年XX月XX日
6. 持有者：Microsoft Corporation
7. **持有者的公钥**
8. 其他
9. 对前面所有项**CA的签名**

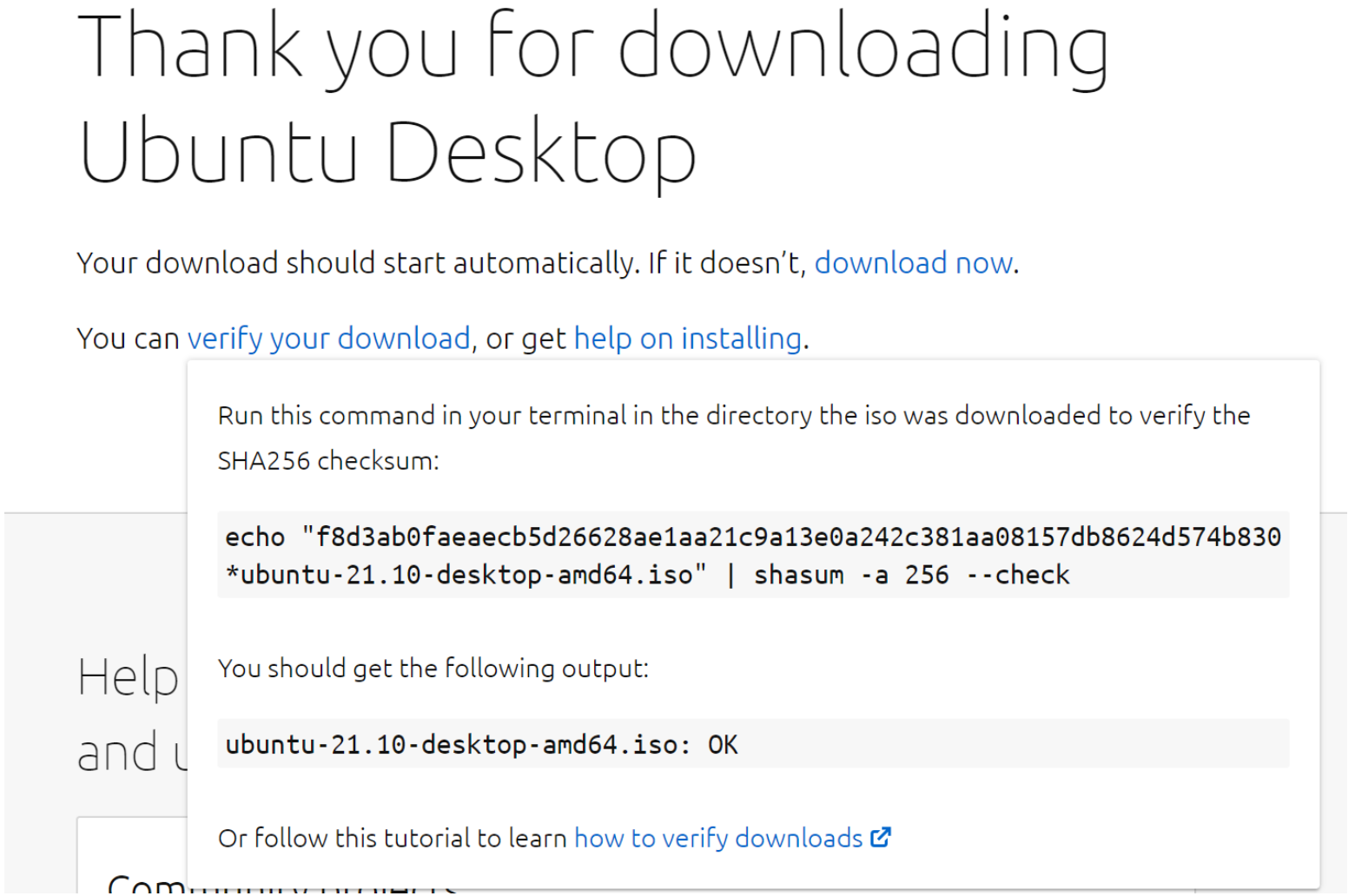
- CA: 可信的认证机构
- CA对数据M签名的过程：
  - h = Hash(M)
  - $S = Sig_{sk}(h)$, sk 是CA的私钥

下载方验证证书

- h = Hash(M)
- $h' = Ver_{pk}(S)$, pk 是CA的公钥
- 验证h, h’是否相同

完整性校验

- 例如下载Ubuntu



一点补充

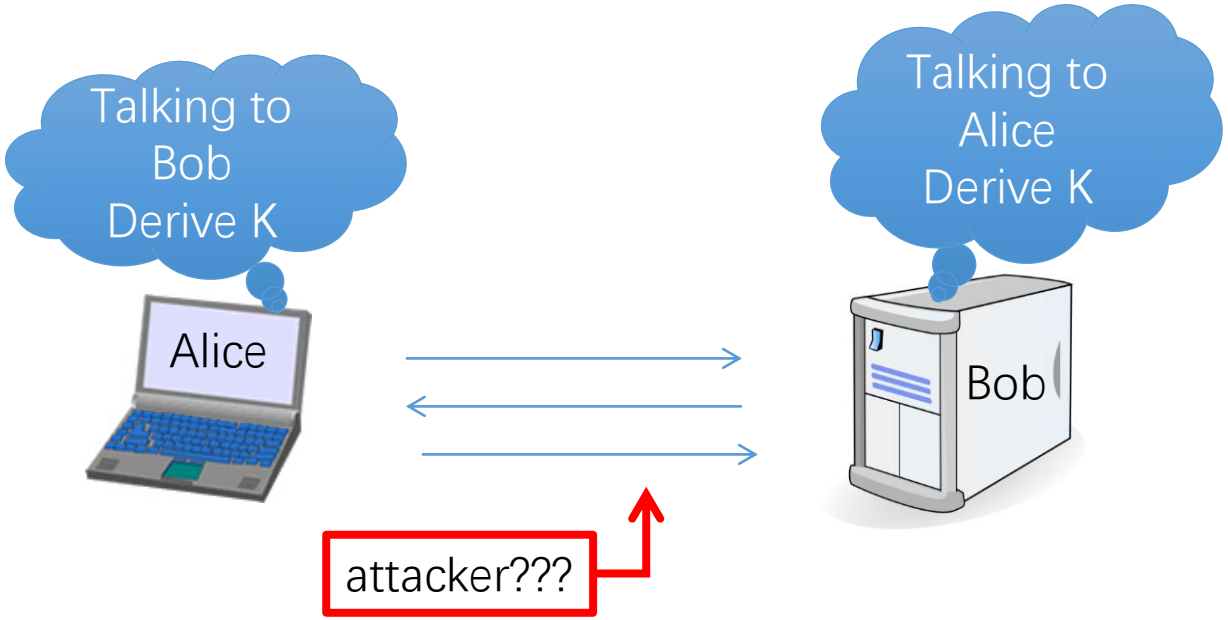
- 密码学是
  - 许多安全机制的基石

- 但是
  - 不能解决所有安全问题
  - 如果实现/使用不当，就变得不可靠
  - 不建议个人自己设计密码算法

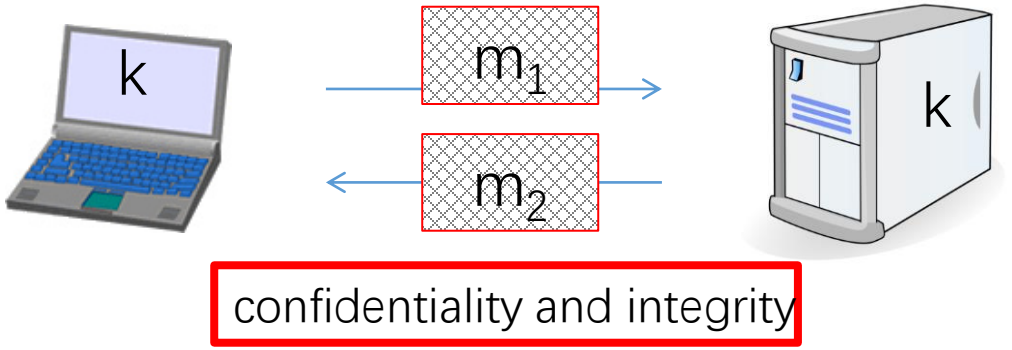
课程简介

密码学的基本功能

密钥协商:

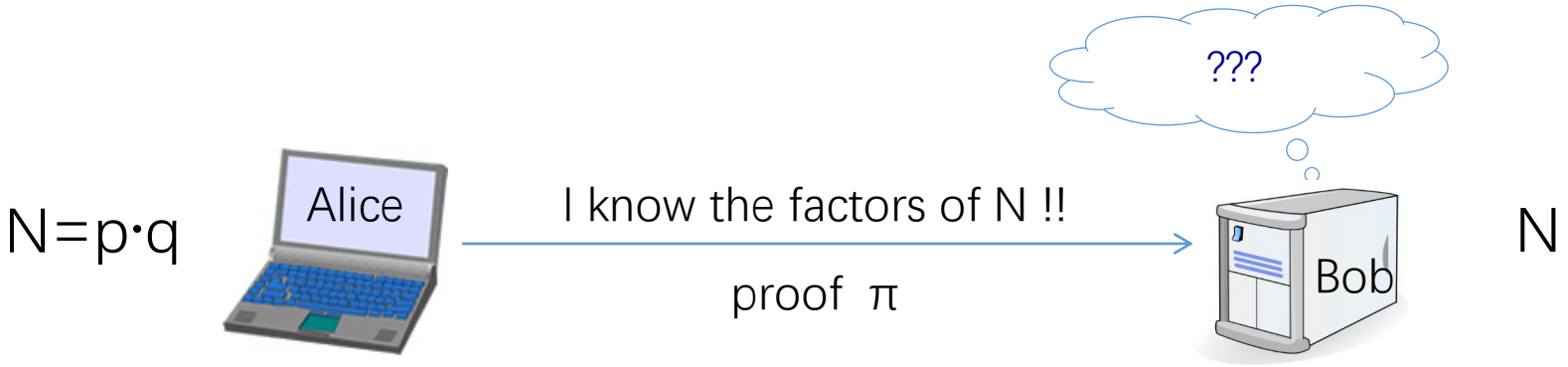


安全数据通信:



更多其他功能

- 数字签名
- 安全计算
- 零知识证明
- ……



课程内容

- 一些对称密码算法、公钥密码算法，及相关的安全性分析

1	引言	9	哈希函数和MD结构
2	安全定义与流密码	10	MD4的碰撞攻击
3	流密码算法与安全性分析	11	基于置换的密码
4	分组密码	12	TMTO攻击
5	差分分析	13	背包与LLL
6	线性分析	14	NTRU
7	AES抗差分/线性攻击自动化分析	15	RSA与Coppersmith方法
8	工作模式和AE		

课程安排

- 时间：
  - 周四 10-11节，即18:30 – 20:10

- 40学时，20周
  - 第1-15周，老师授课
  - 第16-18周，学生汇报考核

课程目标

- 知道一些密码算法、方案的常见漏洞
- 掌握一些经典密码分析方法
- 了解密码算法设计和使用的注意事项

课程需要的基础

- 基础的密码学知识
- 一定的数学基础：基本的线性代数、概率统计、数论等
- 初步的编程能力

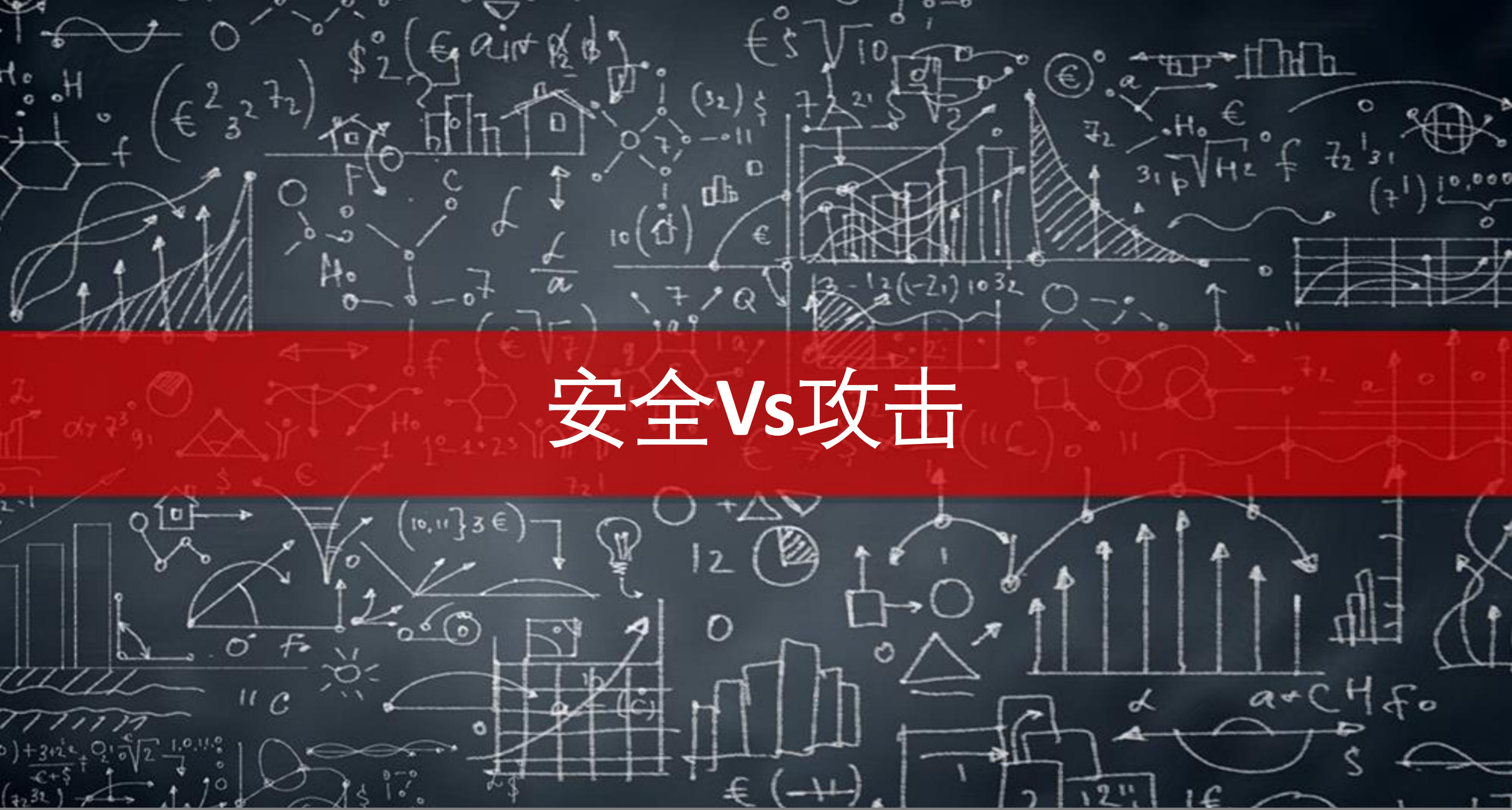
主要参考书

- Mark Stamp, Richard M. Low. **Applied Cryptanalysis: Breaking Ciphers in the Real World.** Wiley-IEEE, 2007.
  - http://www.cs.sjsu.edu/~stamp/crypto/

- Dan Boneh and Victor Shoup. **A Graduate Course in Applied Cryptography**, 2020
  - http://toc.cryptobook.us/

考核方式

- 随堂作业：25%
- 参与度：25%
- Seminar成绩：50%
  - 分组合作
  - 从给定的题目中选一个，研读文献，作报告



安全Vs攻击

密码学的科学性

现代密码学的三部分:

1. 准确定义安全模型
2. 构造具体方案
3. 证明提出的方案符合安全模型

若证明方案不符合安全模型，则构成攻击。

- 结合古典密码，让我们从最朴素的概念开始...
- 安全=?

- 结合古典密码，让我们从最朴素的概念开始...

- 安全=机密性=不让“别人”知道

置换密码

- 将明文字母进行打乱重组的密码
  - 打乱之后的文本就是密文
  - 置换本身就是密钥

Scytale

- 约公元前500年，斯巴达人发明
- 将带子缠绕圆柱形木棍，沿木棍写下文字

T	H	E	T	I	M	E	H	A
S	C	O	M	E	T	H	E	W
A	L	R	U	S	S	A	I	D
T	O	T	A	L	K	O	F	M
A	N	Y	T	H	I	N	G	S

- 展开，沿带子读的文字就是密文：
  - TSATAHCLONEORTYTMUATIESLHMTS…
- 密钥是？破解的难度有多大？

换个方式描述Scytale

- 将字母按行依次写入矩阵，按列读取得到密文
- 例如，对于3*4的矩阵
  - 明文： CRYPTOISFUN

$$\begin{bmatrix} C & R & Y & P \\ T & O & I & S \\ F & U & N & X \end{bmatrix}$$

  - 密文：CTFROUYINPSX
- 密钥是什么？

升级

- 例如
  - 明文：CRYPTOISFUN
  - 矩阵规模 3 x 4，关键字 MATH

$$\begin{array}{c} M\ A\ T\ H \\ \hline \begin{bmatrix} C & R & Y & P \\ T & O & I & S \\ F & U & N & X \end{bmatrix} \end{array}$$

  - 密文：ROUPSXCTFYIN
- 密钥是什么？

如何破解

- 给定如下密文

  VOESA IVENE MRTNL EANGE WTNIM HTMLL ADLTR NISHO DWOEH

- 矩阵规模为m*n
- m, n是什么？
  - 45个字母，所以m*n = 45
- 有多少情况需要考虑？
- 如何知道猜对了？

如何破解

- 给定如下密文

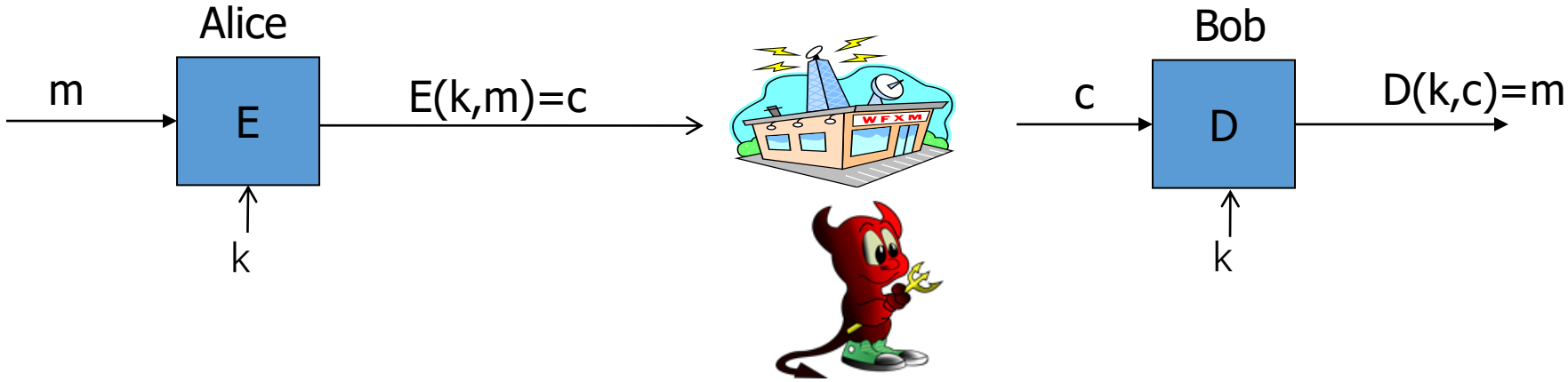
  VOESA IVENE MRTNL EANGE WTNIM HTMLL ADLTR NISHO DWOEH

- 矩阵规模如果是9*5

0	1	2	3	4
V	E	G	M	I
O	M	E	E	S
E	R	W	E	H
S	T	T	A	O
A	N	N	D	D
I	L	I	L	W
V	E	M	T	O
E	A	H	R	E
N	N	T	N	H

2	4	0	1	3
G	I	V	E	M
E	S	O	M	E
W	H	E	R	E
T	O	S	T	A
N	D	A	N	D
I	W	I	L	L
M	O	V	E	T
H	E	E	A	R
T	H	N	N	N

抽象出一个模型



- Alice和Bob进行通信，不希望被窃听
- E，D：加密算法和解密算法
- K：密钥
- m, c：明文，密文
- E, D：可能不公开（现代密码算法基本都公开）
- E, D公开，安全 = 不能由c推断出m来

置换密码的教训

- 置换可以起到一定作用。现代密码学从中演化出“混淆”的概念（Shannon’s principle of diffusion）。
- **穷搜密钥**对攻击者而言永远是一个选择
- 密钥空间不能太小。如果密钥空间足够大，穷搜攻击需要较多时间而令该攻击不现实。
- **足够大的密钥空间是必要的。**
- 但不是充分的。。。

代换密码

- 以英文字母为例，将每一个字母用其他的字母来代替
  - 替换之后的文本就是密文
  - 替换规则就是密钥

凯撒密码 （50 B.C.）

- 明文
  - FOURSCOREANDSEVENYEARSAGO
- 密钥：代换规则

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

- 密文：
  - IRXUVFRUHDAGVHYHABHDUVDIR
- 密钥也即是“平移3个位置”
- 密钥空间多大？

增大密钥空间

- 密钥改成字母的某个置换，不必是平移
- 例如

Plaintext	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
Ciphertext	J	I	C	A	X	S	E	Y	V	D	K	W	B	Q	T	Z	R	H	F	M	P	N	U	L	G	O

- 密钥空间大小 = $26! > 2^{88}$
- 密钥空间足够大

如何破解

- 假如有如下密文

  UKBYBIPOUZBCUFEEBORUKBYBHOBBRFESPVKBWFOFERVNBCVBZPRUBOFERVNBCVBPCYYFVUFOFEIK
  NWFRFIKJNUPWRFIPOUNVNIPUBRNCUKBEFWWFDNCHXCYBOHOPYXPUBNCUBOYNRVNIWNCPOJIOF
  HOPZRVFZIXUBORJRUBZRBCHNCBBONCHRJZSFWNVRJRUBZRPCYZPUKBZPUNVPWPCYVFZIXUPUNFCP
  WRVNBCVBRPYYNUNFCPWWJUKBYBIPOUZBCUIPOUNVNIPUBRNCHOPYXPUBNCUBOYNRVNIWNCPOJ
  IOFHOPZRNCRVNBCUNENVVFZIXUNCHPCYVFZIXUPUNFCPWZPUKBZPUNVR

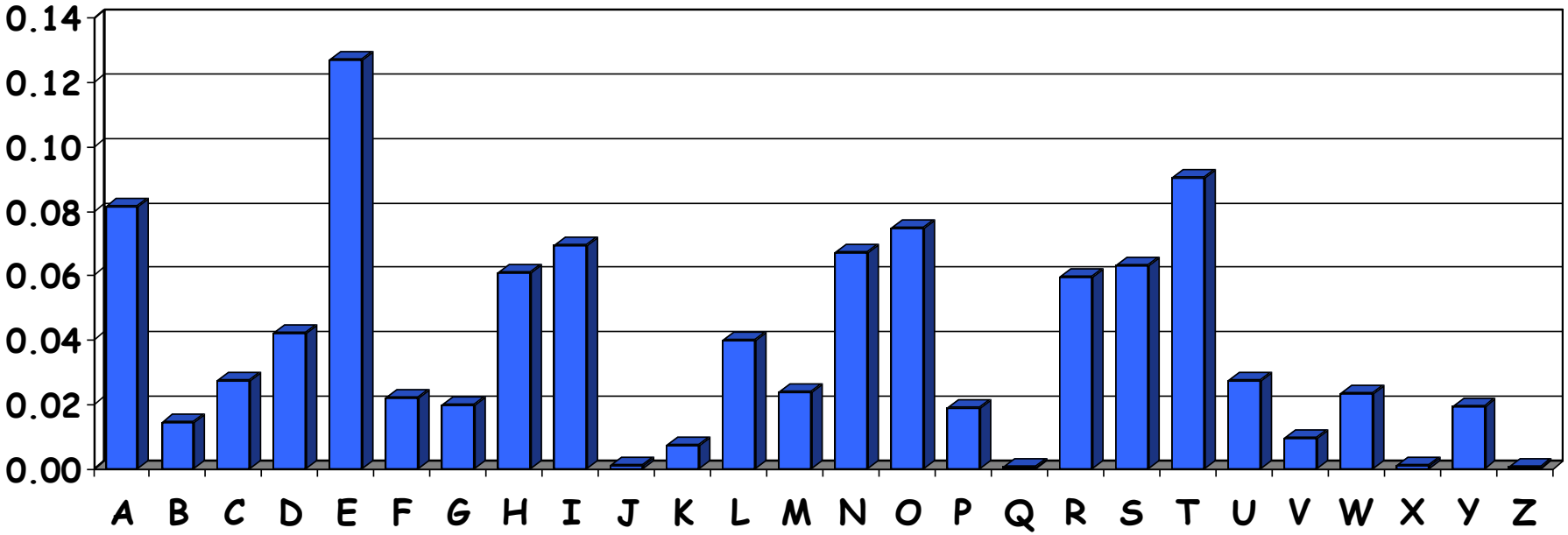
- 穷搜代价太高
- 有没有聪明的方法?
  - 统计规律!

如何破解

- 英语文本中，哪个字母最常见？
  1. X
  2. L
  3. E
  4. H

如何破解

- 英语中不同字母出现的频率



如何破解

- 假如有如下密文

UKBYBIPOUZBCUFEEBORUKBYBHOBBRFESPVKBWFOFERVNBCVBZPRUBOFERVNBCVBPCYYFVUFOFEIKNWFRFIKJN
UPWRFIPOUNVNIPUBRNCUKBEFWWFDNCHXCYBOHOPYXPUBNCUBOYNRVNIWNCPOJIOFHOPZRVFZIXUBORJRUB
ZRBCHNCBBONCHRJZSFWNVRJRUBZRPCYZPUKBZPUNVPWPCYVFZIXUPUNFCPWRVNBCVBRPYYNUNFCPWWJUKB
YBIPOUZBCUIPOUNVNIPUBRNCHOPYXPUBNCUBOYNRVNIWNCPOJIOFHOPZRNCRVNBCUNENVVFZIXUNCHPCYVF
ZIXUPUNFCPWZPUKBZPUNVR

B	36
N	34
U	33
P	32
C	26

➔ E

➔ T

➔ A

NC	11
PU	10
UB	10
UN	9

digrams

➔ IN

➔ AT

UKB	6
RVN	6
FZI	4

trigrams

➔ THE

代换密码的教训

- 代换可以起到一定作用。现代密码学从中演化出“代换”的概念（Shannon’s principle of confusion）。
- **尽管密钥空间足够大，但统计特征可能会暴露明文或密钥的信息。**
- 密文需要看起来**随机**，但随机难以定义也难以达到。
- 统计攻击是一类较难防止的攻击。

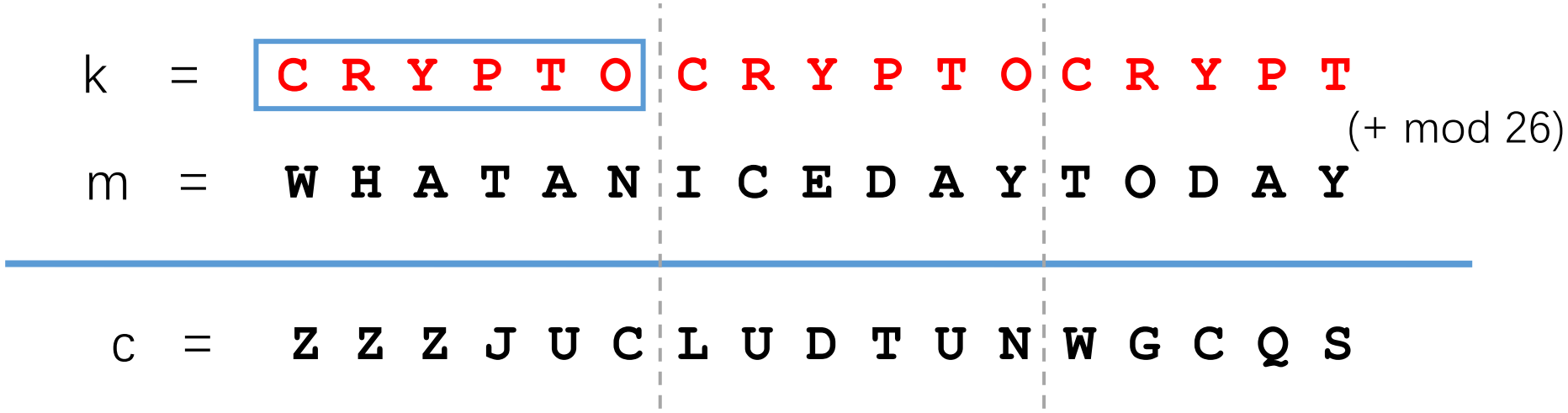
- 如何降低代换密码的统计特征?

- 更具体的问题:

  在上述的单表代换中，字母E被代换成B，B成了密文中的高频字母。
  如何让字母E被代换后频率不那样突出?

  多表代换密码!

多表代换密码——Vigener密码（16世纪，罗马）



- 同一个明文字母，可能被映射到不同的密文字母
- Vigener密码有什么不足？

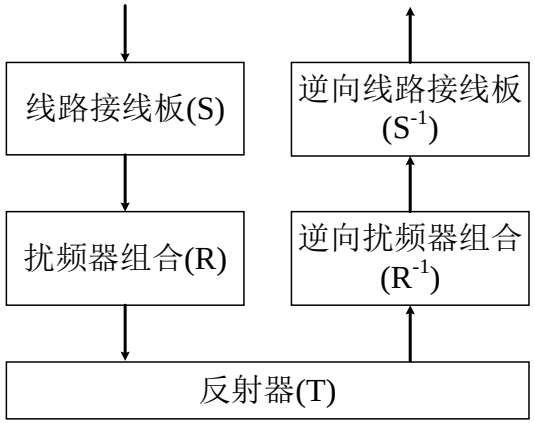
高级多表代换密码——转子密码（1870-1943）

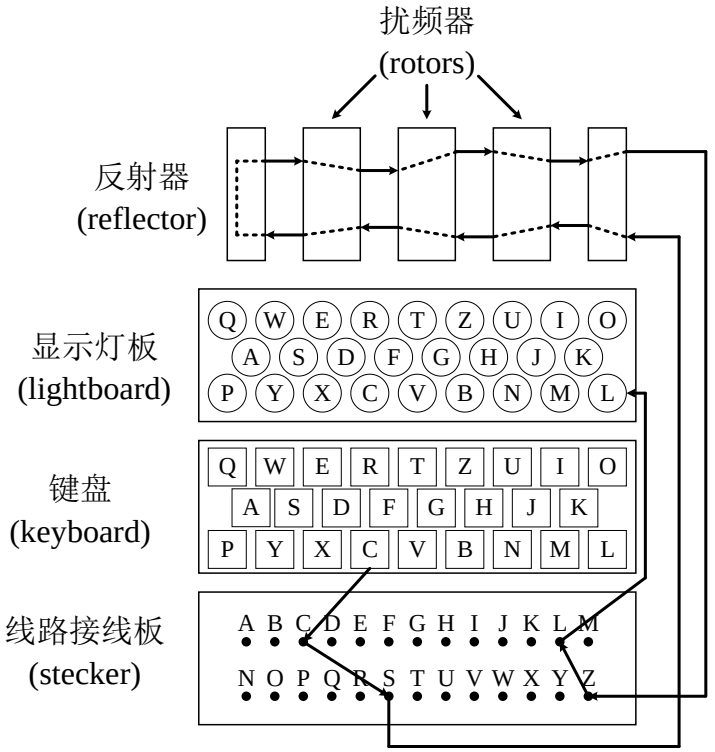
- 最著名的例子：二战德军使用的Enigma
- 5部分组成
  1. 26个字母的键盘，输入组件
  2. **接线板（stecker）**
  3. **扰频组合（rotors）**
  4. **反射器（reflector）**
  5. 显示灯板，输出组件

纳粹德国军用Enigma密码机

高级多表代换密码——转子密码（1870-1943）

- 最著名的例子：二战德军使用的Enigma
- 5部分组成
  1. 26个字母的键盘，输入组件
  2. **接线板（stecker）**
  3. **扰频组合（rotors）**
  4. **反射器（reflector）**
  5. 显示灯板，输出组件



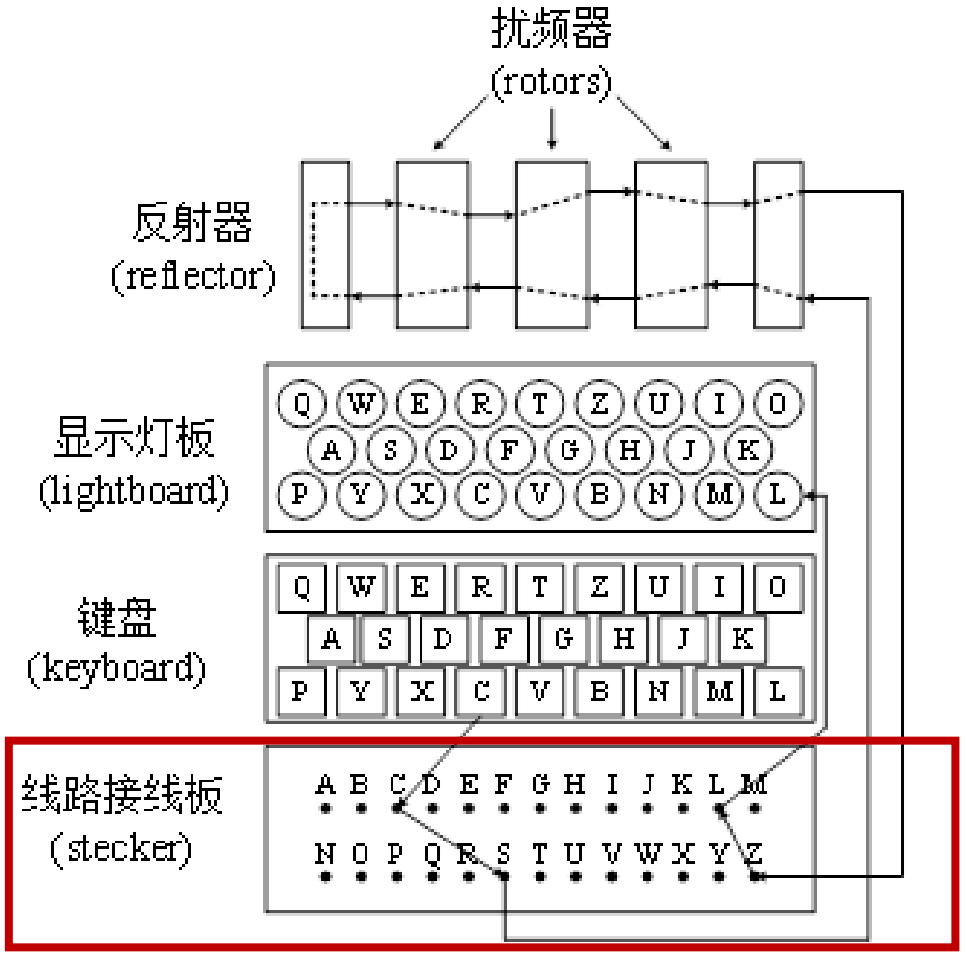


Enigma核心组件介绍

- 线路接线板（S）

  下图接了2条线

输入	**A**	B	C	D	E	F	G	H	I	**J**	K	L	M
输出	**J**	B	C	D	E	F	G	H	I	**A**	K	L	M
输入	N	**O**	P	Q	R	**S**	T	U	V	W	X	Y	Z
输出	N	**S**	P	Q	R	**O**	T	U	V	W	X	Y	Z



Enigma核心组件介绍

- 线路接线板（S）

  下图接了2条线

输入	**A**	B	C	D	E	F	G	H	I	**J**	K	L	M
输出	**J**	B	C	D	E	F	G	H	I	**A**	K	L	M
输入	N	**O**	P	Q	R	**S**	T	U	V	W	X	Y	Z
输出	N	**S**	P	Q	R	**O**	T	U	V	W	X	Y	Z

- 单独看是个单表代换
- 接l条线，一共有多少种代换？

$$f(l) = C_{26}^{2l} \times (2l)!/(l! \times 2^l) = \frac{26!}{(26-2l)! \times l! \times 2^l}$$

- 连接线早期是6条，后增加到10条

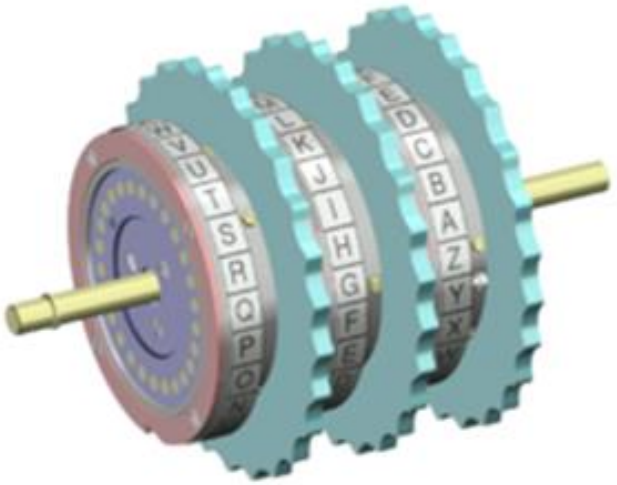
$$f(10) = ?$$

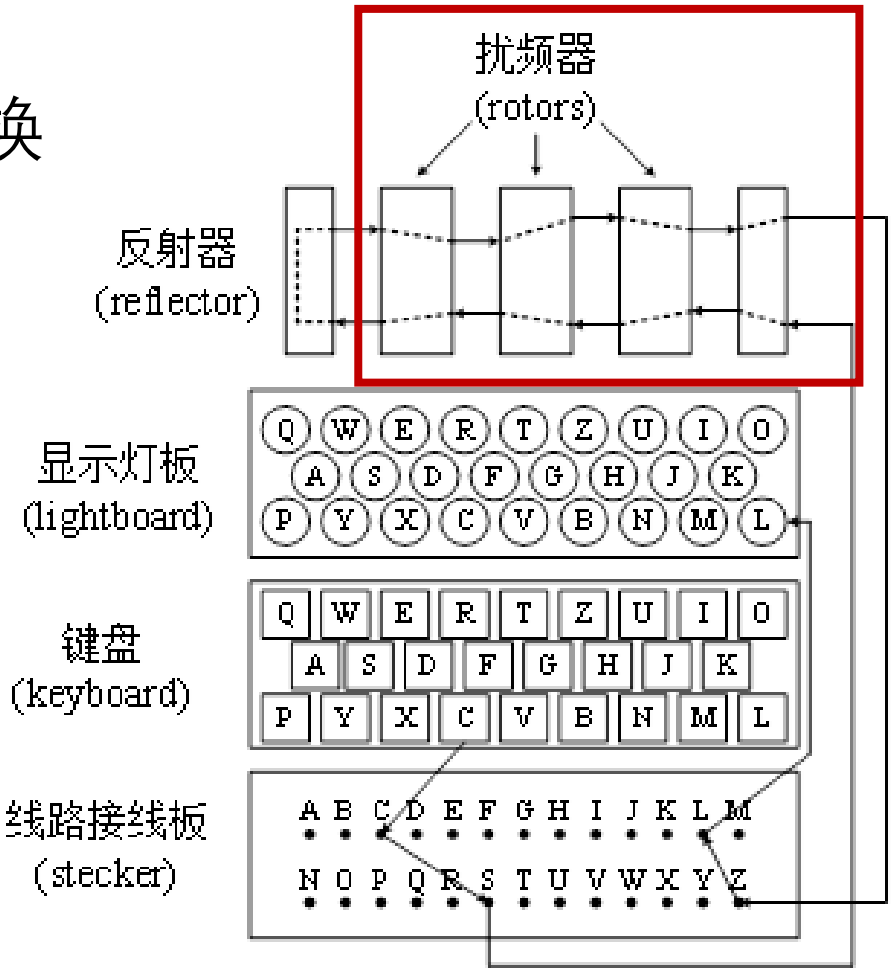
Enigma核心组件介绍

- 扰频器组合（R）

  使用高频、中频、低频三个转子，各为单表代换（类似于秒针，分针，时针）

  - 每加密一个字母，高频转子会旋转一个位置，转了26次后中频转子转一个位置…
  - 扰频器组合形成一个映射，这个映射一直在变，一个字母对应一个代换，**即多表代换**



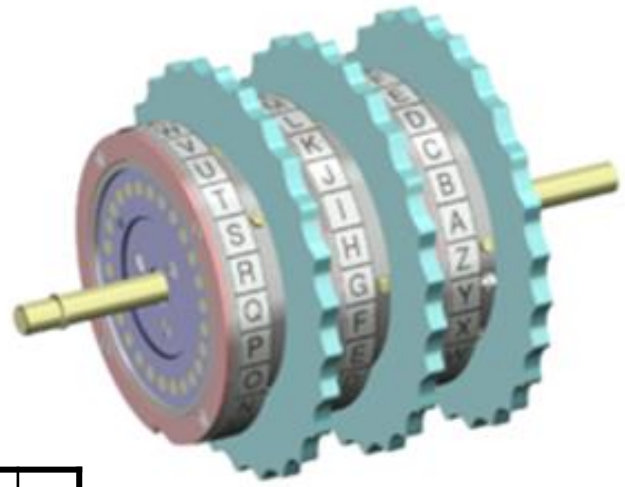


Enigma核心组件介绍

- 扰频器组合（R）

  使用高频、中频、低频三个转子，各为单表代换

  （类似于秒针，分针，时针）



- 扰频器的初始设置有多少种可能性？

  (提示：转子的转动只是平移)

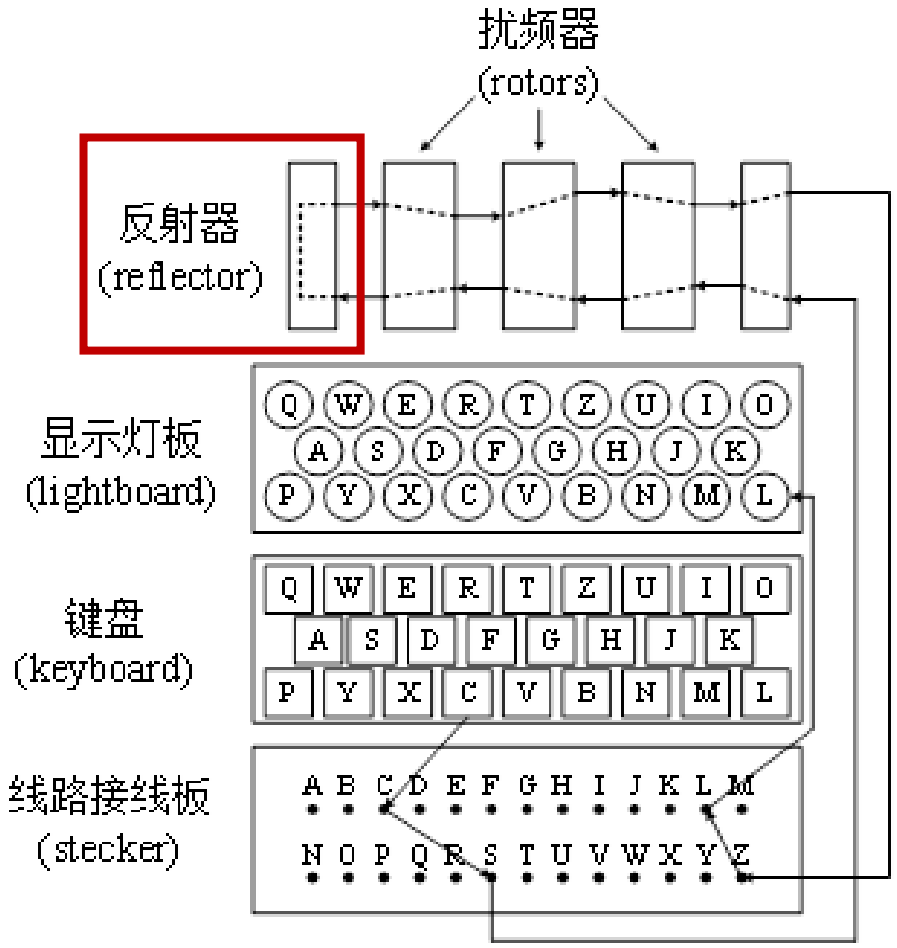
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
初始设置	J	I	C	A	X	S	E	Y	V	D	K	W	B	Q	T	Z	R	H	F	M	P	N	U	L	G	O
下一个	I	C	A	X	S	E	Y	V	D	K	W	B	Q	T	Z	R	H	F	M	P	N	U	L	G	O	J
再下一个	C	A	X	S	E	Y	V	D	K	W	B	Q	T	Z	R	H	F	M	P	N	U	L	G	O	J	I

Enigma核心组件介绍

- 反射器（T）

  固定的单表代换，将**不同**的两个字母进行对换

  字母x经过反射器，出来时不会是x

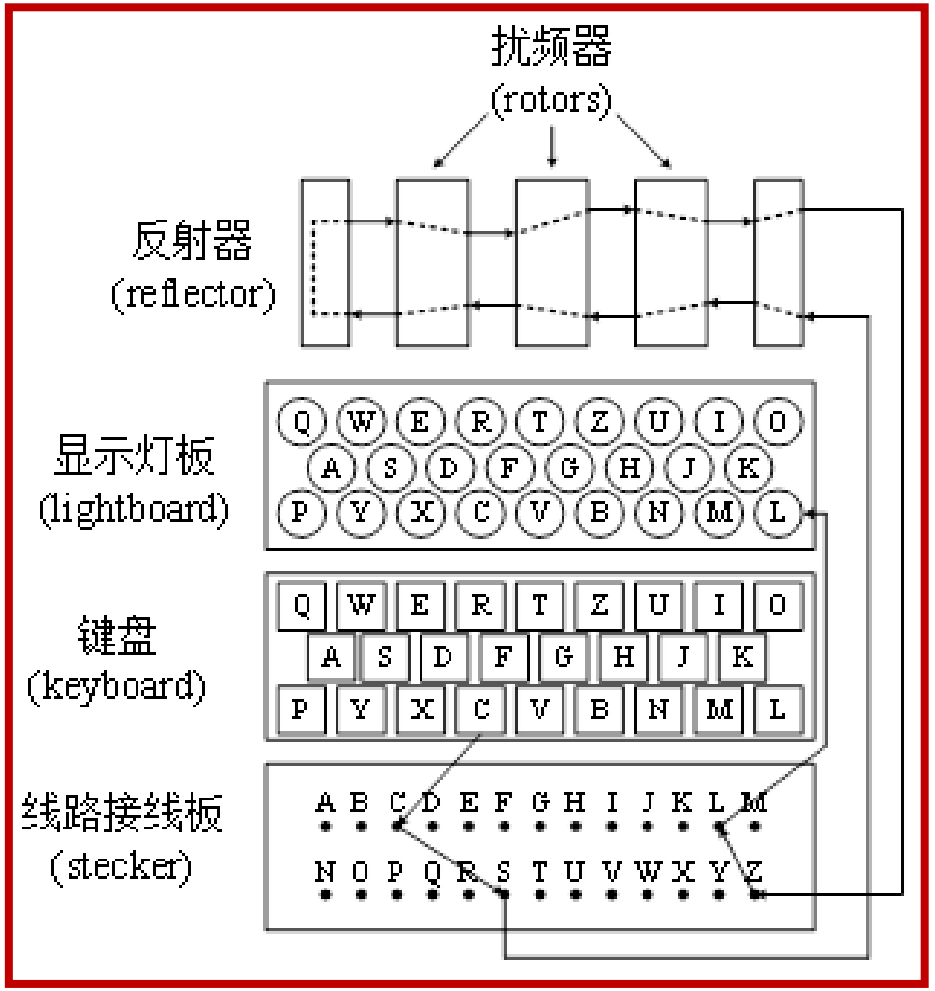


Enigma核心组件介绍

- 看作整体

$$c = S^{-1} \circ R_i^{-1} \circ T \circ R_i \circ S(m)$$
$$m = S^{-1} \circ R_i^{-1} \circ T \circ R_i \circ S(c)$$

- 密钥空间大小
  - 10个连接线
  - 从5个转子中选3个，三个转子的初始位置分别为某个字母朝上
  - 一共有1.59×10^{20}种不同的设置
- 密钥空间足够大；一个字母一个代换，无明显统计特征



如何破解——第一步，找明密文对

- 经验表明，纳粹军队每天早上6点过后会发送一条加密后的规范的天气报告。因此早上6点零5分截获的加密信息几乎必然包含单词WETTER（德语“天气”的意思）。
- Enigma不会将一个字母加密成自身（由于反射器的特点）

猜测的明文 W E T T E R
已知的密文 A E T J W P X E R

（a）✗

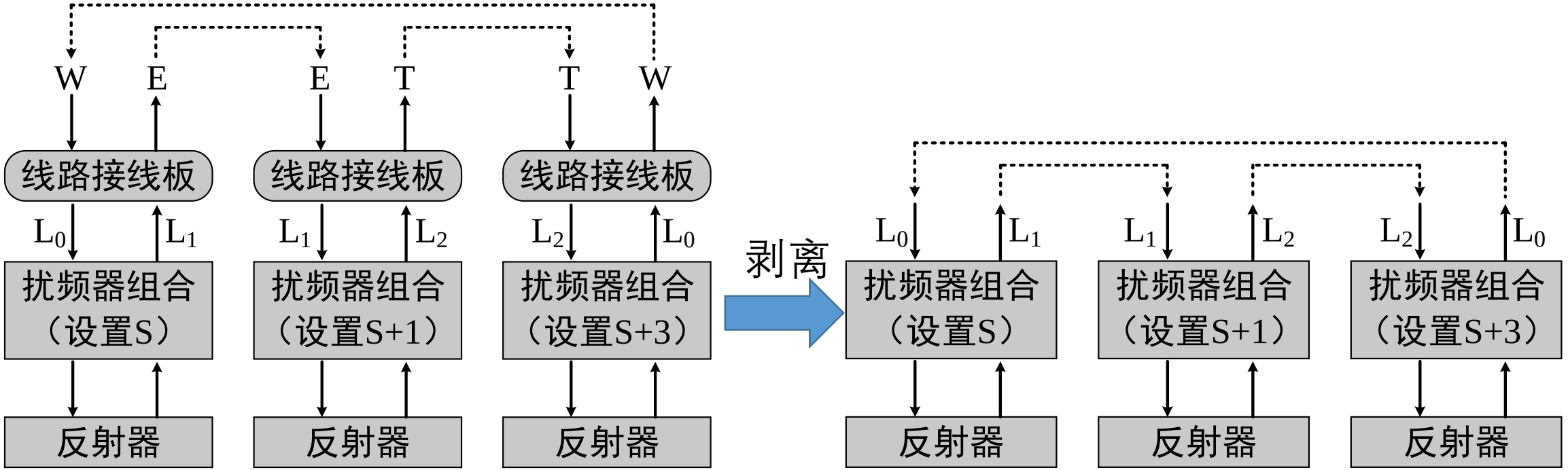
猜测的明文 W E T T E R
已知的密文 A E T J W P X E R

（b）✓

- 根据以上特点，找到可能的 明文-密文 对

如何破解——第二步，如何找到正确的初始设置

- 图灵发现特殊的明密文对 —— 形成环路
- 形成环时，可以先只考虑扰频器和反射器部分，各个击破



1.59×10^{20}种设置

1.05×10^{14}种设置

一点历史故事

- 图灵在1940年初将上述破解思路设计成叫“炸弹（Bombe）”的机械装置，由英国机械工厂负责加工生产。在图灵“炸弹”投入战场使用后，大约需要1个小时就能发现Enigma密码机的设置。
- 使得二战的进程缩短了至少两年，超过一千四百万人的生命得以解救。

高级代换密码的教训

- 避免统计特征非常关键
  - Enigma确实比之前的密码破解难度大多了
- 密码部件之间的关系要足够复杂，避免“**分治攻击**”（**Divide-and-Conquer**）

小结

关于机密性和攻击

- 足够大的密钥空间
- 密文需要看起来随机，避免出现可被利用的统计特征
  - 但随机难以定义也难以达到。
- 密码部件之间的关系要足够复杂，避免“分治攻击”