TECHNISCHE UNIVERSITÄT DRESDEN

PRIVACY AND SECURITY

Stefan Köpsell, Thorsten Strufe

Security and Cryptography 1

Module 2: A little history class…

Disclaimer: Thanks to Dan Boneh, Mark Manulis, Günter Schäfer

Dresden, WS 18

Reprise from Module 1

You know what to expect from the lecture

You have seen some trends that are happening

You have been introduced to some ***typical actors***

You understand what ***threats*** are … and what this means

You can tell ***security goals*** (*CIA!*) from ***security services***

You know ***adversary models*** and which aspects they define

Module Outline

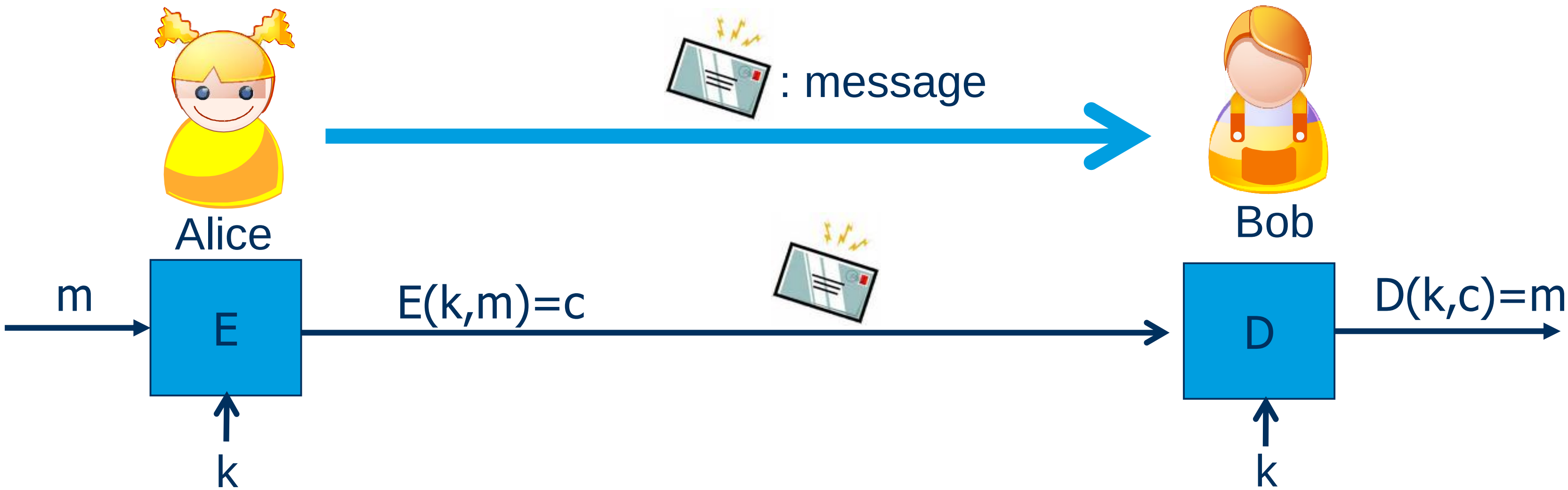
Two words on „Crypto“

A little history of crypto

- Transposition
- Substitution
- Cryptanalysis
- Cesar Cipher
- Vigenère Cipher
- Enigma
- Vernam Cipher – The One Time Pad

Secure Communication

Alice sends Bob a private (any!) message…



m: message (plaintext) $\in$ *M* (message space, sometimes *P*)

k: key $\in$ *K* (key space)

c: ciphertext $\in$ *C* (ciphertext space)

A cipher is a triple of algorithms: **E, D**, *keygen* (sometimes: Enc, Dec)

Correctness: for all $k \in \mathcal{K}$, $m \in \mathcal{M}$: $\mathrm{Dec}(k, \mathrm{Enc}(k, m)) = m$

Terminology: Cryptology

*Crypto**logy**:*

—Science concerned with communications in secure and usually secret form

—Derived from the Greek

- *kryptós (*hidden) and
- *lógos (*word)

—Cryptology encompasses:

- ***Cryptography*** (*gráphein* = to write): principles and techniques by which information can be concealed in *ciphertext* and later revealed by legitimate users employing a secret key

- ***Cryptanalysis*** (*analýein* = to loosen, to untie): recovering information from ciphers without knowledge of the key

Terminology: Cipher

Cipher:

- Method of transforming a message (plaintext) to conceal its meaning (and to transform it back)

- Ciphers are one class of cryptographic algorithms (E,D)
- The transformation usually takes the message and a *(secret) key* as input

- ***Unfortunately***: sometimes also used as synonym for the concealed *ciphertext*

Crypto Basics

Encrypt written communication:

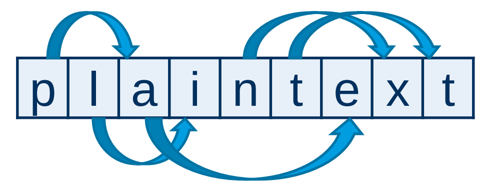
$\mathcal{M}$: language over

$\mathcal{C}$: language over

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
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

$\mathcal{K}$ is determined by a *bijective* mapping

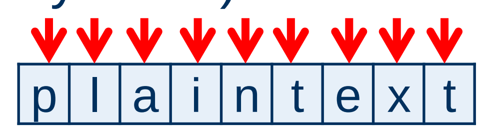
$f : \mathcal{M} \longrightarrow \mathcal{C}$ for Enc and $f^{-1} : \mathcal{C} \longrightarrow \mathcal{M}$ for Dec



Classification

Transposition permute letters according to some scheme

Substitution substitute letters by other letters (or symbols)



A simple Substitution Cipher

PRIVACY AND SECURITY

Key Generation

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
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

choose a *shift value* $k \in [0, 25]$

Encryption

Let $m = m_0 \ldots m_n$ and

Let $\#m_i$ denote the position of m_i in the alphabet.

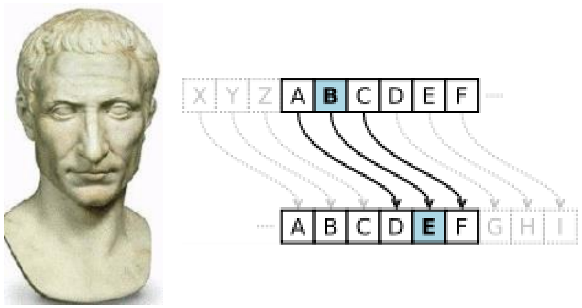
Enc(k, m) = $c_0 \ldots c_n$ where for each c_i : $\mathbf{\#c_i = \#m_i + k \pmod{26}}$

The Caesar Cipher (or: ROT3)

Shift by 3,

$Enc(m_i)$:=

a → D
b → E
c → F
...
y → B
z → C



m = thediehasbeencast

c = WKHGLHKDVEHHQFDVW

What is the size of the key space?

Slightly better Substitution Cipher

Use arbitrary permutation:

k:=

a → X
b → C
c → Y
…
y → V
z → B

More formally:

Key Generation

choose k as a permutation of the alphabet/an l-tuple of distinct symbols

Encryption

Let $m = m_0 \ldots m_n$.

$Enc(k, m) = c_0 \ldots c_n$ where each $c_i = f(k, m_i)$.

What is the size of the key space in this case?

How would you break it?

Breaking a Substitution Cipher

Brute force:

Assuming a shift cipher, try all 26 possible keys

Assuming permutation: 2^{88} (too large)

More intelligent:

1. Use frequency of letters in the expected language
   „e":12.7%, „t": 9.1%, „a": 8.1%, …

Letter Frequencies for Some Languages

Letter	French [3]	German [4]	Spanish [5]	Portuguese [6]	Esperanto [7]	Italian[8]	Turkish	Swedish[9]	Polish[10]	Toki Pona [11]	Dutch [12]
a	7.636%	6.51%	12.53%	14.63%	12.12%	11.74%	11.68%	9.3%	8.0%	17.2%	7.49%
b	0.901%	1.89%	1.42%	1.04%	0.98%	0.92%	2.95%	1.3%	1.3%	0.0%	1.58%
c	3.260%	3.06%	4.68%	3.88%	0.78%	4.5%	0.97%	1.3%	3.8%	0.0%	1.24%
d	3.669%	5.08%	5.86%	4.99%	3.04%	3.73%	4.87%	4.5%	3.0%	0.0%	5.93%
e	14.715%	17.40%	13.68%	12.57%	8.99%	11.79%	9.01%	9.9%	6.9%	7.4%	18.91%
f	1.066%	1.66%	0.69%	1.02%	1.03%	0.95%	0.44%	2.0%	0.1%	0.0%	0.81%
g	0.866%	3.01%	1.01%	1.30%	1.17%	1.64%	1.34%	3.3%	1.0%	0.0%	3.40%
h	0.737%	4.76%	0.70%	1.28%	0.38%	1.54%	1.14%	2.1%	1.0%	0.0%	2.38%
i	7.529%	7.55%	6.25%	6.18%	10.01%	11.28%	8.27%*	5.1%	7.0%	14.8%	6.50%
j	0.545%	0.27%	0.44%	0.40%	3.50%	0.00%	0.01%	0.7%	1.9%	3.0%	1.46%
k	0.049%	1.21%	0.01%	0.02%	4.16%	0.00%	4.71%	3.2%	2.7%	5.1%	2.25%
l	5.456%	3.44%	4.97%	2.78%	6.14%	6.51%	5.75%	5.2%	3.1%	10.2%	3.57%
m	2.968%	2.53%	3.15%	4.74%	2.99%	2.51%	3.74%	3.5%	2.4%	4.4%	2.21%
n	7.095%	9.78%	6.71%	5.05%	7.96%	6.88%	7.23%	8.8%	4.7%	11.6%	10.03%
o	5.378%	2.51%	8.68%	10.73%	8.78%	9.83%	2.45%	4.1%	7.1%	7.7%	6.06%
p	3.021%	0.79%	2.51%	2.52%	2.74%	3.05%	0.79%	1.7%	2.4%	3.7%	1.57%
q	1.362%	0.02%	0.88%	1.20%	0.00%	0.51%	0	0.007%	-	0.0%	0.009%
r	6.553%	7.00%	6.87%	6.53%	5.91%	6.37%	6.95%	8.3%	3.5%	0.0%	6.41%
s	7.948%	7.27%	7.98%	7.81%	6.09%	4.98%	2.95%	6.3%	3.8%	4.1%	3.73%
t	7.244%	6.15%	4.63%	4.74%	5.27%	5.62%	3.09%	8.7%	2.4%	4.6%	6.79%
u	6.311%	4.35%	3.93%	4.63%	3.18%	3.01%	3.43%	1.8%	1.8%	3.2%	1.99%
v	1.628%	0.67%	0.90%	1.67%	1.90%	2.10%	0.98%	2.4%	-	0.0%	2.85%
w	0.114%	1.89%	0.02%	0.01%	0.00%	0.00%	0	0.03%	3.6%	2.8%	1.52%
x	0.387%	0.03%	0.22%	0.21%	0.00%	0.00%	0	0.1%	-	0.0%	0.04%
y	0.308%	0.04%	0.90%	0.01%	0.00%	0.00%	3.37%	0.6%	3.2%	0.0%	0.035%
z	0.136%	1.13%	0.52%	0.47%	0.50%	0.49%	1.50%	0.02%	5.1%	0.0%	1.39%

most frequent

2nd most frequent

Di- and Trigram Frequencies

Digram (Bigram) Frequencies

— frequency for a combination of two letters

English: th, he, in, en, nt, re, er, an, ti, es, on, at, se, nd, or, ar, al,…

German: er, en, ch, te, nd, ei, de, ie, in, es, ge, ne, un, ic, st, an,…

Trigram Frequencies

— frequency for a combination of three letters

English: the, and, tha, ent, ing, ion, tio, for, nde, has, nce, edt, tis,…

German: ein, ich, der, sch, und, die, nde, cht, ine, den, end, che, ens,…

A small Example

Given the ciphertext C:

LIKHKDGDQBWKLQJFRQILGHQWLDOWRVDBKHZURWHLWLQFLSKHUWKD
WLVEBVRFKDQJLQJWKHRUGHURIWKHOHWWHUVRIWKHDOSKDEHWWKD
WQRWDZRUGFRXOGEHPDGHRXWLIDQBRQHZLVKHVWRGHFLSKHUWKHVH
DQGJHWDWWKHLUPHDQLQJKHPXVWVXEVWLWXWHWKHIRXUWKOHWWH
URIWKHDOSKDEHWQDPHOBGIRUDDQGVRZLWKWKHRWKHUV

35 W	e,t,a?
33 H	e,t,a?
24 K	e,t,a?
20 D	e,t,a?
18 R	
16 L	
15 Q	

10 KH	he,..,th?
5 WK	he,..,th?
4 KD	
3 WW	

bigrams

2 WWK	
2 WKH	the,and?
2 RIW	the,and?
2 HWW	
1 ZRU	the,and?

trigrams

$ grep -o . file.txt | sort -f | uniq -ic | sort -rg

A small Example – solved

PRIVACY AND SECURITY

```
$ rotix -f enc_cesar.txt -r23
ifhehadanythingconfidentialtosayhewroteitinciphertha
tisbysochangingtheorderofthelettersofthealphabetthat
notawordcouldbemadeoutifanyonewishestodecipherthesea
ndgetattheirmeaningshemustsubstitutethefourthletterof
thealphabetnamelydforaandsowiththeothers
```

If he had anything confidential to say, he wrote it in cipher, that is, by so changing the order of the letters of the alphabet, that not a word could be made out. If anyone wishes to decipher these, and get at their meaning, he must substitute the fourth letter of the alphabet, namely D, for A, and so with the others.

– Suetonius, Life of Julius Caesar

Ciphertext-only attack!

Breaking a Substitution Cipher

Brute force:

Assuming a shift cipher, try all 26 possible keys

Assuming permutation: 2^{88} (too large)

More intelligent:

1. Use frequency of letters in the expected language
„e“:12.7%, „t“: 9.1%, „a“: 8.1%, …
2. Use frequency of bi-grams, tri-grams…

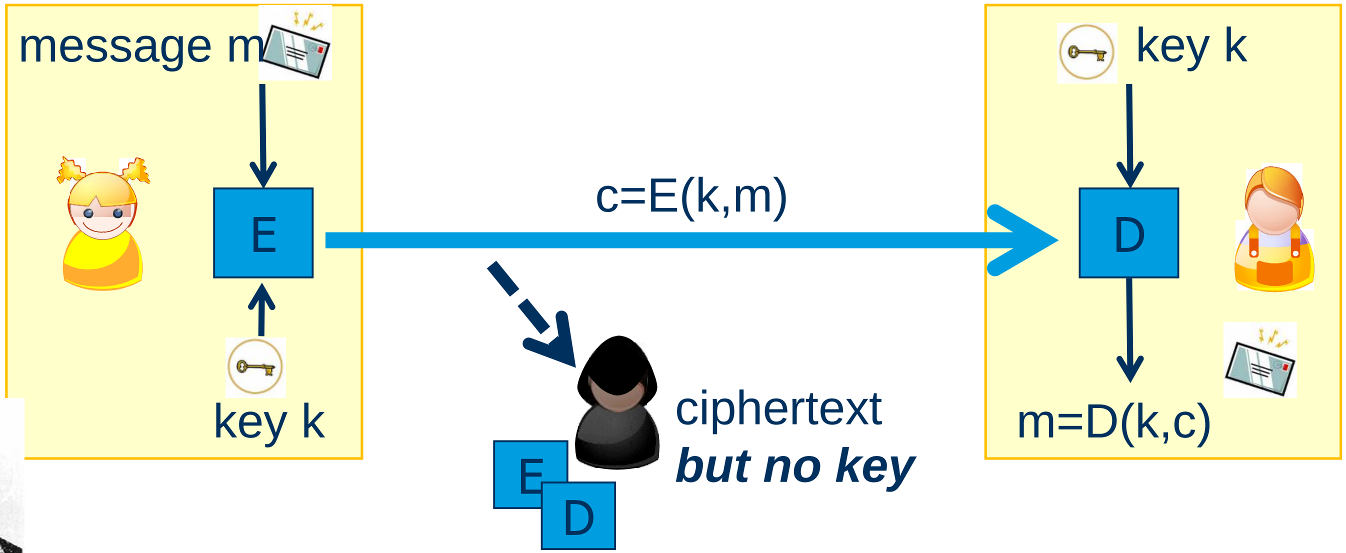
Would you know an immediate remedy to make cipher more secure?

Auguste Kerckhoffs (1835 – 1903)

1. *The system should be, if not theoretically unbreakable, unbreakable in practice.*
2. *The design of a system should not require secrecy, and compromise of the system should not inconvenience the correspondents.*
   (Kerckhoffs' principle)
3. *The key should be memorable without notes and should be easily changeable.*
4. *The cryptograms should be transmittable by telegraph.*
5. *The apparatus or documents should be portable and operable by a single person.*
6. *The system should be easy, neither requiring knowledge of a long list of rules nor involving mental strain.*

Auguste Kerckhoffs: „ La Cryptographie Militaire“ in „le Journal des Sciences Militaires“, 1883

The Communication Model and Kerckhoff



"The cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience."

i.o.w: KGen, E, and D will inevitably be discovered at some stage

→ All algorithms should be public

→ security must rely on secrecy of the key only

Polyalphabetic Substitution (Vigenère)

Monoalphabetic cipher easily broken by statistics

Goal: decrease impact of language statistics

What could you do?

Concept:

1. use periodic, variable substitution
2. define (and communicate) periods by key

Key Generation

choose key $k = (k_1, \ldots, k_d)$ where each k_i is defined through *shift value*

Encryption

Let $m = m_0 \ldots m_n$.

$Enc(k, m) = c_0 \ldots c_n$ where $\mathbf{c_i = f(k_{i+1}, m_i)}$ and index of k is taken mod d.

Vigenère Poly-alphabetic Table

2: Substitute as usual..

1: Chose alphabet from keyword

Recta transpositionis tabula.

a b c d e f g h i k l m n o p q r s t u x y z w
b c d e f g h i k l m n o p q r s t u x y z w a
c d e f g h i k l m n o p q r s t u x y z w a b
d e f g h i k l m n o p q r s t u x y z w a b c
e f g h i k l m n o p q r s t u x y z w a b c d
f g h i k l m n o p q r s t u x y z w a b c d e
g h i k l m n o p q r s t u x y z w a b c d e f
h i k l m n o p q r s t u x y z w a b c d e f g
i k l m n o p q r s t u x y z w a b c d e f g h
k l m n o p q r s t u x y z w a b c d e f g h i
l m n o p q r s t u x y z w a b c d e f g h i k
m n o p q r s t u x y z w a b c d e f g h i k l
n o p q r s t u x y z w a b c d e f g h i k l m
o p q r s t u x y z w a b c d e f g h i k l m n
p q r s t u x y z w a b c d e f g h i k l m n o
q r s t u x y z w a b c d e f g h i k l m
r s t u x y z w a b c d e f g h i k l m n
s t u x y z w a b c d e f g h i k l m n o
t u x y z w a b c d e f g h i k l m n o p
u x y z w a b c d e f g h i k l m n o p q
x y z w a b c d e f g h i k l m n o p q r
y z w a b c d e f g h i k l m n o p q r s
z w a b c d e f g h i k l m n o p q r s t
w a b c d e f g h i k l m n o p q r s t u

In hac tabula literarū canonica siue recta tot ex uno & usuali
latinarum literarum ipsarum permutationem seu transposition
alphabeta, quot in ea per totum sunt monogrammata, uidelice
& uigesies quatuor & uiginti, quæ faciunt in numero D.lxxvi.
tidē multiplicata, paulo efficiunt minus q̄ quatuordecē milia.

Vigenère Cipher

k	=	C	R	Y	P	T	O	C	R	Y	P	T	O	C	R	Y	P	T	
																			(+ mod 26)
m	=	W	H	A	T	A	N	I	C	E	D	A	Y	T	O	D	A	Y	
c	=	Z	Z	Z	J	U	C	L	U	D	T	U	N	W	G	C	Q	S	

How would you break it?

Suppose most common $|k|^{th}$ letter = „D“

Test: first letter of keyword = „D“ – „E“ = Y…

Try incremental keyword length

Kasiski‘s Differences

PRIVACY AND SECURITY

Assuming a ciphertext:

T	I	G	I	M	K	Z	O	T	I	G	V	M	C	Z	O	A	O	F	W	L	J	Z	D	E	M	X	H	X	M	X	L	G	J	O	H	R	Y	C	P	L	M	C	W	X	F	I	R	T	I	G	F	M	Y	E	H	T	U	Y	H	P	A	K

Find repeating patterns and their distance:

TIG : 1,9,49 (-> distance 8, 40)

ZO : 7,15 (-> distance 8)

	Distance	Potential key length	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
TIG	40		X		X				X												X
ZO	8		X		X				X												

(subsequently: analyse frequencies…)

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

Better Poly-Alphabetic Encryption

Observation: short keys can be guessed easier

Concept: Choose key

1. as long as the message (no repitions)
2. with highly varying letters

m =	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
k =	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
c =	M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M

→ Key does not repeat, key contains many different characters!

Can you still break it? How?

PRIVACY AND SECURITY

t	h	e												o	x	j							y	z	l								
t	h	e												t	h	e							t	h	e								
M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M

			u	n			k	b				t	f						p	s					l	l			a	z			
			i	n			i	n				i	n						i	n					i	n			i	n			
M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M

t	h	e	u	n			k	b				t	f	o	x	j			p	s			y	z	l	l			a	z			
t	h	e	i	n			i	n				i	n	t	h	e			i	n			t	h	e	n			i	n			
M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M

					e	d				o	w										k	l					e	l					
					i	s				i	s										i	s					i	s					
M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M

t	h	e	*u*	*n*	*e*	*d*	*k*	b		o	w	***t***	f	o	x	j	u	m	p	s	*k*	l	y	z	l	l	e	l	a	z				
t	h	e	*i*	*n*	*i*	*s*	*i*	n		i	s	***io***	n	t	h	e	S	E	i	n	*l*	s	t	h	e	n	i	s	i	n				
M	O	I	C	A	M	V	S	O	X	W	O	B	S	H	E	N	M	Q	X	F	S	D	R	G	T	Y	M	D	I	M	K	D	M	

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

Product Ciphers

Observation

Statistics of language and periodicity cause weakness

Goal: Make predictions again harder

Product Ciphers:

combine different transformation and/or substitution ciphers:

let $F_1, \ldots, F_d$ be different ciphers (with same spaces $\mathcal{K}$, $\mathcal{M}$, $\mathcal{C}$)

$$Enc(k, m) = F_d \circ \ldots \circ F_2 \circ F_1(m)$$

Rotor Machines

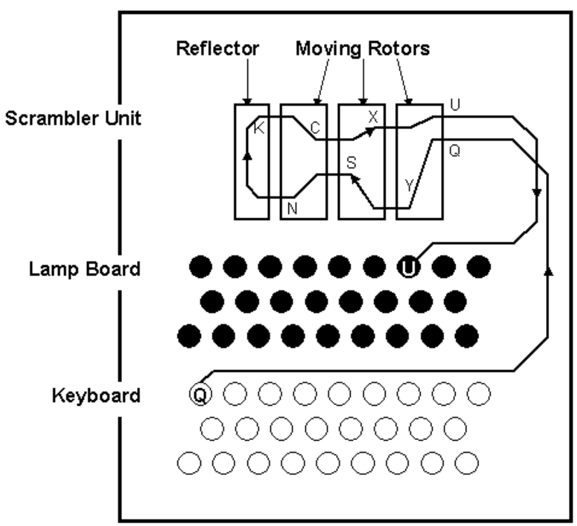
Idea: change alphabet independent of key (extend periods)

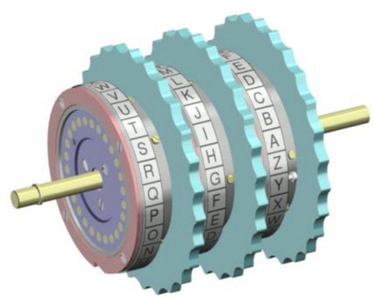
rotors $R_1, \ldots, R_d$; each R_i performs simple substitution

Rotors rotate incrementally after each encrypted character

wirings ensure *polyalphabetic* substitution (~26^d keys)

reflector allows Enc & Dec to be the same algorithms





The Enigma
Used in WWII; Visit Bletcheley Park!

Breaking Enigma

Brute force, ciphertext only: too many possible keys

Observation: No letter is ever encoded to itself

C	O	H	J	Y	P	D	O	M	Q	N	J	C	O	S	G	A	W	H	L	E	I	H	Y	S	O	P	J	S	M	N	U
Pos 1			K	E	I	N	E	B	E	S	O	N	D	E	R	E	N	E	R	**E**	**I**	G	N	I	S	S	E				
Pos 2				K	E	I	N	E	B	E	S	O	N	D	E	R	E	N	E	R	E	I	G	N	I	S	S	E			
Pos 3					K	E	I	N	E	B	E	S	**O**	N	D	E	R	E	N	**E**	R	E	I	G	N	I	S	**S**	E		

Positions 1 and 3 for the possible plaintext are impossible because of matching letters. The red cells represent these *crashes*. Position 2 is a possibility.

© Wikipedia

Goal: reduce possible solutions

Guesses:

— codeword changed infrequently

— frequent similar plaintexts

Idea: guess plaintext, vary position

Marian Rejewski (1905-1980)

Alan Turing (1912-1954)

Perfect Secrecy

Observation: Patterns are your enemy!

Concept:

- Long key (long/no periodicity)
- No recognizable pattern

Gilbert Vernam
(1890-1960)

Key Generation

choose $k = (k_1, \ldots, k_n)$ where each k_i is truly random permutation

Encryption

Let $m = m_0 \ldots m_n$.

$Enc(k, m) = c_0 \ldots c_n$ where $c_i = f(k_i, m_i)$

The One-Time-Pad (Vernam cipher)

Cannot be broken with CTO, not even brute force attack…

Truly random key, as long as the message:

m_0 =	A	T	T	A	C	K	T	H	E	C	I	T	Y	A	T	T	W	E	L	V	E	(+ mod 26)
k =	P	S	P	I	U	H	G	D	S	P	H	G	D	S	P	I	W	E	E	W	O	
c =	P	L	I	I	W	R	Z	K	W	R	P	Z	B	S	I	B	S	I	P	R	S	
k =	Y	H	P	R	S	R	G	F	F	D	D	X	N	S	Q	I	S	P	W	N	F	(+ mod 26)
m_1 =	R	E	T	R	E	A	T	F	R	O	M	C	O	A	S	T	A	T	T	E	N	

… or any other message of the same length, for that matter

Now, what are the two problems with this method?

Summary

PRIVACY
AND
SECURITY

You‘ve learned the tuple (M,C,K,E,D)

You can tell the difference of

- Cryptology
- Cryptography
- Cryptanalysis

You know basic cryptanalysis (statistics/combinatorics)

You know transposition and substitution

You have learned the shift cipher

… Mono- and Polyalphabetic substition ciphers

… Product ciphers

You know CTO attacks

You understand why we model security as a game!

And you have heard of the legends of Enigma and Alan Turing…