History of Cryptography

Simon Singh: The Code Book

Contents

1. **Hand encryption**
   - Monoalphabetic codes
   - Semialphabetic codes
2. **Encryption machines**
   - Encryption disks
   - Enigma
3. **Computer Encryption**
   - Standard encryption protocols
   - Public Key Encryption
   - Quantum Cryptography

Cryptography

Cryptology (= the science of secrecy) consists of

- cryptography = secret writing
- cryptanalysis = analysis, detection of the secret

The history of cryptology is a constant battle between cryptography and cryptanalysis. Many discoveries have been disclosed after a long delay, as part of a military strategy, because they gave an advantage to one side in the war.

We will use the terms

- plaintext - original message (we write it in lower letters)
- ciphertext - encrypted message (we write it in capital letters)

For a plaintext we use 26 letters of the English alphabet.

Transposition codes

Transposition codes reorders letters in the plaintext.

- "Along a fence" - we write letters of the plaintext alternately to the left and right along the fence; the ciphertext starts with all letters on the left side of the fence and continues with all leters on the right side.
- "Belt around a stick" - wrap a belt around the stick and write the plaintext on the belt in several lines (rotate the stick and continue the text along the stick). The unrolled belt then contains the ciphertext with the letters rearranged.
  To decrypt, you need to use a stick of the same diametr.

Substitutional monoalphabetic codes

Substitution codes replace the letters of the plaintext with other letters or characters.

Caesar shift code

- Caesar, 100-44 BC, Roman Empire.
- Each letter of the plaintext is shifted by the difference d, $p \to p + d$.
- Example: The distance $d = 3$

plaintext	v e n i. v i d i. v i c i.
ciphertext	Y H Q L. Y L G L. Y L F L.

- Decrypting: Try all 26 options for the key d.

Substitutional monoalphabetic codes

Monoalphabetic code

- Any one-to-one representation between letters of the plain alphabet to letters of the cipher alphabet may be used.
- Number of bijections between alphabets is $26! \doteq 4 \cdot 10^{26}$. This can be no longer done by brute force without a computer.

Kerckhoffs principle

- Kerckhoffs, 1885, Netherlands
- The security of an encryption system must not depend on the secrecy of the algorithm, but only on the secrecy of the key.

Substitutional monoalphabetic codes

Monoalphabetic code

- Practical implementation: write the key phrase at the beginning of the cipher alphabet without spaces and repeating letters and add the remaining letters "alphabetically".
- Example: The key phrase is Julius Caesar.

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

plaintext	v e n i. v i d i. v i c i.
ciphertext	M S Y R. M R I R. M R L R.

Substitutional monoalphabetic codes

Frequency analysis

- Deciphering of monoalphabetic codes was brought by the development of mathematics and linguistics in Arabia, the 9th century. (The 12th century in Europe).
- Frequency analysis of the text - the frequency of occurrence of letters in the text of a given language is different. A frequent character in a ciphertext is likely to mean a frequent letter in that language.

English:	e - 12.7%,	t - 9.1%,	a - 8.2%;
	q - 0.1%,	z - 0.1%	
Czech:	e - 10.9%,	a - 9.6%,	o - 8.0%;
	w - 0.05%,	x - 0.03%,	q - 0.005%

- For messages shorter than 100 letters frequency analysis fails.

Substitutional monoalphabetic codes

Advanced monoalphabetic codes

- To place clauses (in order to decieve) to different places in the text, to use characters for all words.
  That didn't work - see the execution of Mary Stuart, 1587, England
- Syllable-by-syllable encryption
  The great code - father and son Rossignol, 1650-1680, France.
  Their cipher alphabet had 587 different characters (triplets).
  The letters of Kings Louis XIII and Louis XIV resisted to decryption for another 200 years.
- The homophonic substitution code, the 17th century Europe.

Substitutional monoalphabetic codes

Homophonic substitution code

- Each letter is assigned as many characters (doubles) as its average frequency in the language. This gives an average frequency of 1% for each character of the ciphertext.
- It is a monoalphabetic code because the cipher alphabet is the same throughout the text (the same ciphertext character corresponds to the same plaintext letter).
- Frequency analysis for letter pairs can be used to decipher it.
  English has frequent combinations: ee, th, qu
  The letter q (0.1% frequency) is followed only by the letter u (2.8% frequency). In the ciphertext, the single character for q will be followed by one of the three characters for u.

Substitutional polyalphabetic codes

Polyalphabetic codes

- Leon Battista Alberti, 1460, Italy.
  He proposed a regular alternation of two cipher alphabets.
- Blaise de Vigenère, Treatise on codes, 1586, France.
  He proposes alternating multiple cipher alphabets according to an agreed key.
- Vigenèr code seemed difficult to use for his contemporaries. It took 200 years to catch on, and it spread with the invention of the telegraph, 1851. (This is also when Morse code was invented.)

Substitutional polyalphabetic codes

Vigenèr code

- Vigenèr code uses the Vigenèr square of all shifted cipher alphabets with distance d, where $1 \leq d \leq 26$.
- The keyword is repeatedly written over the plaintext. The letter "P" of the keyword indicates the shift of the corresponding letter of the plaintext - the alphabet in the line starting with "P" is used to encrypt it.
- Example. The key word is CHLEB.

KlÄ‰ovŠ slovo	c	h	l	e	b	c	h	l	e	b	c	h
Posun o d	2	7	11	4	1	2	7	11	4	1	2	7
OtevĹen text	v	e	n	i.	v	i	d	i.	v	i	c	i.
Ĺifrov text	X	L	Y	M.	W	K	K	T.	Z	J	E	P.

Substitutional polyalphabetic codes

Vigenèr square

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	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
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
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
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
⋮																									
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
⋮																									
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
⋮																									
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

Substitutional polyalphabetic codes

Breaking through the Vigenèr code

- Charles Babbage, 1854, England (during the Crimean War); published by Friedrich Wilhelm Kasiski, 1863, Prussia.
- We note the repeated words in the ciphertext and their distances. If this are longer words (at least four letters), it is likely that they correspond to the same plain word and that they were encrypted with the same cipher alphabet. Their distances v_i are multiples of the length of the keyword, we guess a length of the keyword $l = \gcd(v_1, \ldots, v_k)$.
- We split the text into l subtexts encrypted with the same alphabet and use frequency analysis.

Substitutional polyalphabetic codes

Until the World War I

- Invention of radio, Marconi, 1894 - easy communication and also easy eavesdropping. Big push to develop cryptography.
- Many codes were invented during the World War I, but they were all variations on an old theme and were broken soon. (Zimmermann's telegram to Mexico about the German attack on the US.)
- One-time-pad code
  Designed by Joseph Mauborgne, 1918, US.
  Patented by Gilbert Vernam, 1919, Bell Laboratories, US.

Substitutional polyalphabetic codes

One-time-pad code

- Uses a random key of the same length as the plaintext, the key determines the alphabet selection (= shift for a given letter as in the Vigenèr code).
- The key must be truly random, if it containes meaningful words there is room for frequency analysis in it.
- Both the sender and the receiver must have the same book of random keys, and each key can only be used once.

Substitutional polyalphabetic codes

One-time-pad code

- Vernam patented this code for binary words. A shift of 0 or 1 is done by $\oplus$ (logical *XOR*), or the sum modulo 2.
- Encryption and decryption using the key k is done the same way: $a \to a \oplus k = b$; $b \to b \oplus k = a$.
- If we use one key twice, we get some information about the plain messages from the cipher messages: $(a_1 \oplus k) \oplus (a_2 \oplus k) = a_1 \oplus a_2$

Substitutional polyalphabetic codes

One-time-pad code

The one-time-pad code is *unbreakable!*

Even if we would be able to try all the keys (there are 26^d of them, where d is the length of the message), we would get all the meaningful texts of length d. But we don't know which one of them has been sent!

Practical difficulties:

- producing keys - they must be really random
- distribution of keys - each must be used only once

Code is used for the hotline between Washington and Moscow.

Encryption Machines

- Encryption disks have been used for encryption since the 15th century (Alberti, 1460, Italy).
- Two disks are placed on a common axis so that they can be rotated independently. The outer disk containes a plain alphabet, the inner disk a cipher alphabet.
- Discs are suitable for both the Caesar shift code and the Vigenèr code (after encrypting one letter one should rotate the disk so that the next letter of a password is under the letter A of the plain alphabet).

Enigma

- The Enigma machine, which was used by German army during the World War II, is the most famous encryption machine at all.
- Enigma was invented by a businessman Arthur Scherbius, in 1918 in Germany. It was a kind of "typewriter" measuring 34x28x15 cm and weighing 12 kg.
- Because of its high price, neither the military nor the business community was interested in it at first. The German army started to use it later on in 1926.

Enigma

Description of Enigma

- Enigma consisted of a keyboard for a plaintext, three rotary disks (scramblers) and a signal board for a ciphertext.
- A plate (reflector) was placed at the back side to return a signal through scramblers back by a different route, which allowed to place the signal board above the keyboard.
- The connection of the wires from the keyboard through the scramblers to the bulbs of the signal board indicates the bijection of a plain alphabet to a cipher alphabet.
- The plaintext was typed on the keyboard and it put on the light at the ciphertext letters on the signal board.

Enigma

Description of Enigma

- After typing one letter, the first scrambler rotates "one letter" forward, which changes the cipher alphabet.
- Scramblers are used as "ones, tens, hundreds" - after the first scrambler is rotated 360 degrees, the second scrambler is rotated in 1/26 from 360 degrees, then the first scrambler is rotated again, etc.
- This is a polyalphabetic code with $26^3 = 17\,576$ cipher alphabets. This guarantees resistance to the frequency analysis.

Enigma

Description of Enigma

- The initial setting of scramblers (i.e. what letter is at the top of them) gives a key for encryption. However, the number of 17 576 keys is not big enough.
- Scramblers are removable and their order can be swapped, for three scramblers there are $3! = 6$ possibilities of reordering.
- Between the keyboard and the scramblers is a solid linking board that connects the twelve letters into six pairs. This does not increase the number of cipher alphabets since the board is fixed, but the number of keys will increase $\frac{1}{6!}\prod_{i=0}^{5}\binom{26-2i}{2} = 100\ 391\ 791\ 500$ times.
- The total number of keys is roughly 10^{16}, which guarantees resistance to the brute force breaking.

Enigma

Encryption and decryption key

- The reflector board causes the same setting for both, encryption and decryption, both use the same key.
- A daily encryption/decryption key consists of a scrambler setting (e.g. H-Q-L), an order of scramblers (e.g. 2-3-1) and pairwise connections (e.g. A/L - P/R - T/D - B/W - K/F - Q/Y).
- The daily key is used to encrypt a current message key (which encludes scrambler settings only, e.g. D-Y-G). Then the scramblers are reconfigured for the message encryption.
- Daily keys were distributed in the German army once a month in so-called monthly books.

Enigma

Enigma

Battle with Enigma - Poland

- Poland felt threatened by Germany, so it worked hard to break Enigma even before World War II (1932-39).
- Thanks to espionage, they got the device and its documents. The idea was to break the daily key.
- Marian Rejewski took advantage of the fact that Germans repeated a currant key twice before a message (e.g. D-Y-G-D-Y-G). The key was encrypted with the same daily key. A repetition is a gap for a jeweler.

Enigma

Battle with Enigma - Poland

- Rejewski processed many messages encrypted with the same daily key. From the three pairs of repeated characters at their beginnings he created strings of cipher characters (cycles in the bijection of cipher alphabets for the 1st and 4th letter, etc.) and noticed that their lengths are determined only by the order and setting of the scramblers.
- He worked for a year to explore all $6 \cdot 26^3 = 105\,456$ possibilities and compiled a catalogue of "string lengths".
- With the catalog, he was ready to decrypt. He assembled the strings of present day characters, found the appropriate order in the catalog, and set up the scramblers for the daily key. He tried decrypting a text without the linking board, and then it was easy to guess the six linked letter pairs.

Enigma

Battle with Enigma - Poland

- Later on Rejewski built a machine that mechanically searched for scramblers settings for given strings of characters. Six machines with different scramblers order worked in parallel and found a corresponding settings in about two hours.
- In 1938, the Germans increased the security of Enigma by making two new types of scramblers - selecting 3 out of 5 possible scramblers, which is $5 \cdot 4 \cdot 3 = 60$ possibilities - and linking 20 letters into 10 pairs. The Poles didn't have the money to build that many decrypting machines.
- In 1939, July 24, the Poles passed their results to French and English cryptographers.

Enigma

Battle with Enigma - England

- Alan Turing was looking for a method of decryption that didn't depend on repeating a currant key twice. He used strings of characters made up of "hints" (e.g. at 6:05 the Germans used to sent a weather report beginning with WETTER).
- Turing completed his work in March 1940, the Germans stopped repeating currant keys in May. Turing also perfected decryption machines (called bombs because they ticked while decrypting) till August and it took only one hour then to break the daily key.
- The English continued to decrypt Enigma even after the war, since other countries such as the USSR used it till 1950s. England publicly revealed its results in 1974.

World War II encryption

- In addition to Enigma, the Germans used the Lorenz code for communication betwen Hitler and his top commanders. The English also it broke too. The Japanese used the Purpur code.
- The Americans did not have an invented code, but used the language of the Navajo, a native Indian tribe, whose language structure is completely different from any European or Asian language.
- The code talkers soldiers were Navajo. They were able to memorize a huge number of new words (e.g. fighter plane = hummingbird, bombs = eggs). An encryption and decryption was then done in real time of a phone call.
- The Navajo language was one of the few codes in history that was never broken. It was only declassified by the US government in 1968.

Computer encryption

- Decryption machines were the predecessors of computers (Colossus machine for decrypting the Lorenz code, built from electron tubes, 1943, England).
- The invention of the transistor in 1947 made it possible to make computers wider. IBM began making computers in 1953. In 1959, the integrated circuit was invented. In the 1960s more and more companies owned computers and used them for encryption.
- Letters are encoded in ASCII code as binary words (American Standard Code for Information Interchange, IBM, US, 1960-67). Classical codes can be used for encryption at a bit level, i.e. within a letter. For example, a transposition code that swaps two adjacent bits, the Vernam code. Various numerical operations can also be performed on binary words.

Computer encryption

Standard encryption protocols

- In 1973, the American Bureau of Standards announced a competition for a standard encryption protocol for the US. The leading candidate was the Lucifer protocol, designed by Horts Feistel of IBM.
- The NSA (=National Security Agency) weakened this protocol to 56-bit keys, and this version was officially adopted in 1976 under the name DES (=Data Encryption Standard).
- DES works with binary words of length 64, splits them into two parts of half the length, runs one part through a function and adds the other part and repeats this sixteen times.

Computer encryption

Standard encryption protocols

- AES (=Advanced Encryption Standard) has been used since 2002. AES works with words of length 128, which it treats as a matrix of bytes of size 4×4. This matrix is "compiled" by a sequence of operations using a $128-$bit key.
- Both DES and AES are symmetric encryption protocols that use the same key for encryption and decryption. The main problem with using these protocols is key distribution. Before a secret communication can occur, both parties must share one piece of secret information, namely the key.

Computer encryption

Key distribution problem

- Whitfield Diffie, Martin Hellman, Ralph Merkle, US, tried to solve the key distribution problem.
- Diffie predicted the emergence of the internet and the need for encryption that involved a general public. Distributing keys by agents with briefcases would then be impossible.
- The vault with two locks ("mine-mine and yours-yours") proves that secret information can be exchanged without first exchanging keys. But encryption functions do not usualy commute with decryption functions!
- Hellman, 1976, proposed a public secret key agreement protocol using a discrete logarithm. A discrete exponential function is a one-way function; composing two exponential functions commutes (as "mixing collors").

Computer encryption

Key distribution problem

- Diffie, 1975, invented a public key encryption. It uses asymmetric keys - a public key for encryption and a secret key for decryption - and a one-way function with a backdoor for encryption. The value of it can be computed quickly, but without the private key it is difficult to invert it.
- Advantages - no synchronous communication is required when arranging the key. The public key is still available in the "phone book".
  Everyone has a single key to communicate with everyone else.
  Private key can be used to digitally sign messages.
- The paper was a revolution in cryptology, but Diffie did not have a concrete proposal for a public-key protocol.

Key distribution problem

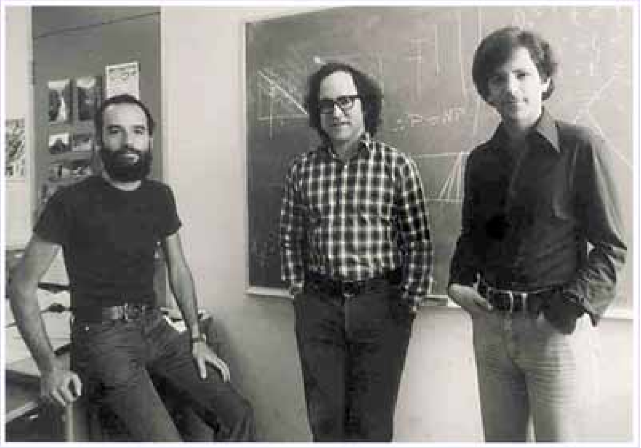
Ralph Merkle, Martin Hellman, Whitfield Diffie

Computer encryption

Public key cryptography

- Ronald Rivest, Adi Shamir, Leonard Adleman, US, patented the first public key encryption protocol.
- The RSA protocol uses the factorization problem. The one-way function is a discrete power, and the Euler-Fermat theorem opens the back door.
- The paper was published in Scientific American in 1977, along with a competition to decrypt with the modulus $n = pq \doteq 10^{129}$ for a 100 dolars reward. The modulus n was factorized after seventeen years by the combined efforts of many computers.

Public key encryption



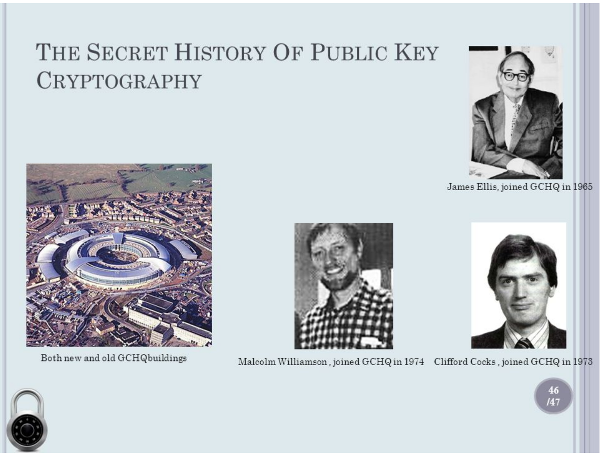
Adi Shamir, Ronald Rivest, Leonard Adleman

Computer encryption

Public key encryption

- James Ellis, Clifford Cocks, Malcolm Williamson, England, members of GCHQ, invented the same thing, but their research was top secret.
- Ellis, 1969 - the idea of public key encryption (so-called unclassified encryption).
- Cocks, 1973 - found a one-way function relying on the factorization problem (RSA protocol).
- Williamson, 1975 - found a public secret key agreement relying on the discrete logarithm problem (Diffie-Hellman protocol).
- These results were not published until 1997.

Public key encryption



THE SECRET HISTORY OF PUBLIC KEY CRYPTOGRAPHY

James Ellis, joined GCHQ in 1965

Both new and old GCHQbuildings

Malcolm Williamson , joined GCHQ in 1974

Clifford Cocks , joined GCHQ in 1973

46 /47

James Ellis, Malcolm Williamson, Clifford Cocks

Computer encryption

Pretty good privacy

- Phil Zimmermann, US, 1991, published PC software called PGP (=Pretty Good Privacy), which uses RSA encryption to exchange a symmetric key and then it encrypts an actual messages faster with the symmetric IDEA code.
- PGP software can also generate keys for RSA and can attach a digital signature to messages.
- Zimmermann put PGP on the web for free download and was charged with illegal arms export. He was later acquitted.
- The question is: Does a citizen have a right to privacy or does the state have a right to eavesdrop to ensure security?

Quantum encryption

Quantum computers

- If quantum computers can be made, the security of all known encryption protocols falls.
- David Deutsch, 1984 - the first quantum computer idea.
- Peter Shor, 1994 - quantum factorization in polynomial time (the end of RSA security).
- Lov Grover, 1996 - quantum list search in polynomial time (searches all keys for AES).

Quantum encryption

Quantum key distribution

- Stephen Wiesner, Charles Bennett, Gilles Brassard, 1984, US, invented quantum key agreement.
- By measuring the polarization of photons, one can arrange for an arbitrarily long secret key. Moreover, any eavesdropping is recognized because it changes the polarization of the photons.
- The key can then be used for a One-time-pad code, which (in the case of a random key) is unbreakable.
- In 1995, it was possible to implement quantum key agreement using optical fibre in Geneva at a distance of 23 km.

The History of Cryptography

Literature

- Simon Singh: The Code Book. Random House, 2000.