

A Very Brief History of Encryption

Part 1 Caesar to Enigma

Larry Wittig

Lexington Computer and Technology Group

March 2012

Overview of this presentation

This presentation mostly follows The Code Book by Simon Singh:

- Simple translation and substitution encryption
 - Vigenere (polyalphabetic) encryption to Enigma (WWII)
 - The advent of electronic computers and the Arpanet/Internet
 - Diffie-Hellman Exchange
 - RSA (public key – private key stuff)
 - Pretty Good Privacy (PGP)
 - SSL & TLS (the **s** in **https**)
- The older stuff
- The newer stuff – since the advent of electronic computers

Besides The Code Book I also consulted Wikipedia (quite often) and other Web pages. I generally tried to give the links.

Discussed in Singh's book but not covered here in any detail:

- Mary Queen of Scots vs. QE1
- Navajo code talkers during WW II
- Beale treasure papers
- Quantum computers
- Deciphering lost languages and ancient scripts

History of encryption pre Electronic Computers

- Earliest known encryption dates back to Egypt about 2000 BCE, and a simple **translational** encryption scheme that was used by Julius Caesar is well documented.
- **Substitution** (random) ciphers were still used to the late middle ages when they were replaced by **polyalphabetic** ciphers. They were used for both personal and state secrets.
- Before and during World War II, mechanical and **electromechanical polyalphabetic** machines were in wide use (e.g. the German Enigma machine).

There are thousands of other “non-computer” or “paper and pencil” ciphers. These are just the most important ones.

Simple Shift/Translational Substitution Encryption

- A SHIFT of 3 letters is known as a Caesar shift cipher.

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

- HELLO becomes KHOOR
- A variation on this is to shift the alphabet by a KEYWORD with the rest of the letters translated but without reusing any of the letters in the KEYWORD.



























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

- Capitalization, punctuation and the space between words is generally omitted, or the encrypted text may group the letters in sets of N characters.
- Note that on a very elementary level, an eavesdropper (Eve) could know the cipher scheme but would find decoding difficult without the keyword.
- For more on the Caesar cipher see: http://en.wikipedia.org/wiki/Caesar_cipher

Substitution Encryption

(beyond simple translation)


- In this case a random letter or a symbol is substituted for each letter:





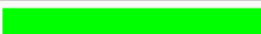

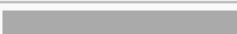
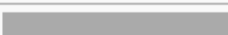

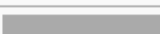
















A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
D	M	F	V	N	B	X	I	E	L	T	A	G	C	H	Y	U	O	W	K	Z	R	S	J	Q	P
																									

- See: <http://www.mathsatwhitehaven.com/mathsclub/codes/codeindex.htm#extools>
- Substitution encryptions can be easily broken with “frequency analysis” especially if the message is long. This was known by Arab scholars as early as 800 (C.E.)
- Homophonic cyphers use more than one code letter/symbol per plaintext letter as a way to minimize the effectiveness of frequency analysis
- Substitution encryptions can be made even more complicated by substituting for pairs of letters or for syllables. (See Singh on the Great Cipher on Louis XIV.)

Substitution Decryption

Frequency analysis

- The longer the encrypted message the easier it is to decrypt.
- For more on frequency analysis see: http://en.wikipedia.org/wiki/Letter_frequency 
- For a web site that works you thru a decryption example see <http://www.esg.montana.edu/meg/consbio/cryptogram/crypto.html>
- For information on common pairs, triplets and repeated letters see <http://www.counton.org/explorer/codebreaking/frequency-analysis.php>

Letter ↕	Relative frequency in the English language ▾	
e	12.702%	
t	9.056%	
a	8.167%	
o	7.507%	
i	6.966%	
n	6.749%	
s	6.327%	
h	6.094%	
r	5.987%	
d	4.253%	
l	4.025%	
c	2.782%	
u	2.758%	
m	2.406%	
w	2.360%	
f	2.228%	
g	2.015%	
y	1.974%	
p	1.929%	
b	1.492%	
v	0.978%	
k	0.772%	
j	0.153%	
x	0.150%	
q	0.095%	
z	0.074%	

Polyalphabetic Cipher

a.k.a. Vigenere Cipher

- Singh: the Vigenere cipher [was named] in honor of the man who developed it into its final form in the 16th century.
- Like substitution cipher except the substitution alphabet changes after each of the first N characters, and then the scheme repeats. It makes simple frequency analysis more difficult.
- **REQUIRES A KEYWORD** – this is an important new and enduring aspect, so security of the keyword becomes extremely important – the Vigenere square scheme could be public and secrecy depended solely on the keyword
- See the excel worksheet in a program called Code Frequencies about 90% of the way down
<http://www.mathsatwhitehaven.com/mathclub/codes/codeindex.htm#extools>
- Around 1850 Charles Babbage showed that frequency analysis was still able to solve polyalphabetic ciphers if they were of sufficient length. However, his method was kept private probably at the request of British Intelligence.

Cipher disk

(As opposed to a hard disk drive that's been encrypted)

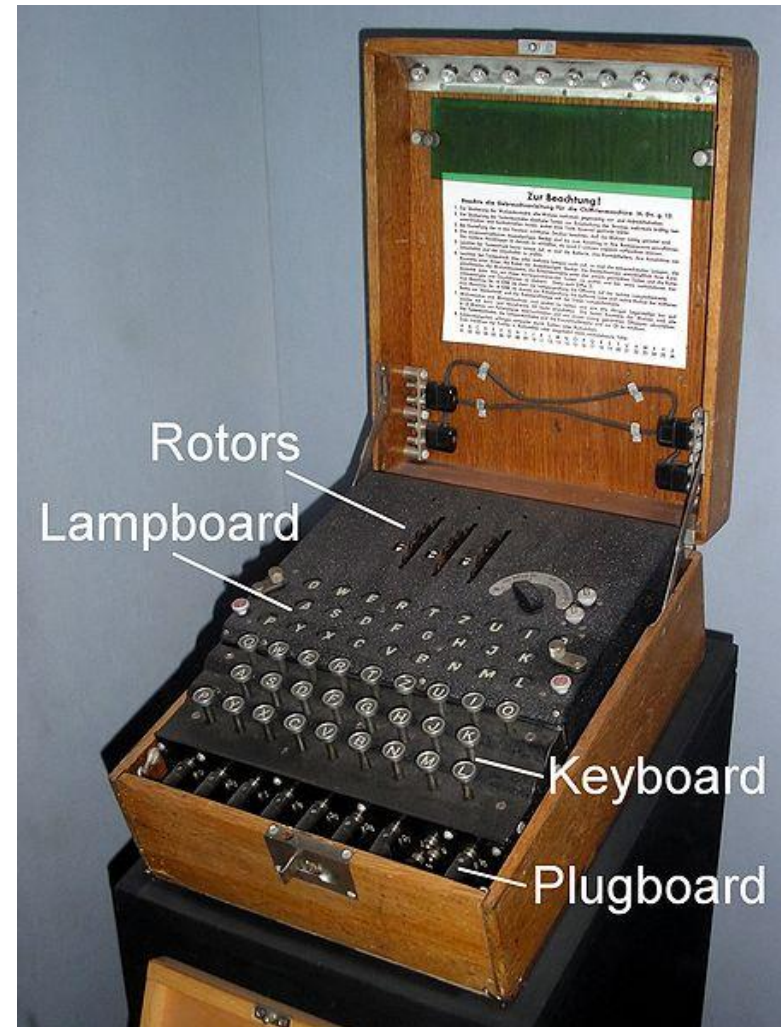
- Described by Alberti in 1467
- The inner disk and outer ring could rotate w.r.t. each other.
- The letters on the inner disk could be in alphabetical order for a **simple translation cipher**.
- More generally they could be scrambled for a **substitution cipher**.
- Alice and Bob would both set their disks up by aligning a given inner disk character with a given outer ring character – this means they have to exchange some (keyword-like) information ahead of time.
- If the disks were rotated after each character by some agreed upon keyword scheme it could be used to handle **polyalphabetic encryption**.



The cypher disk shown here is dated 1893, and is similar ones were used in the US civil war.

German Enigma Machine

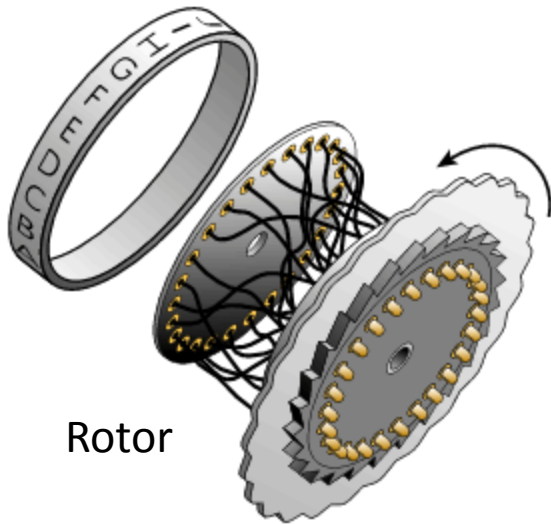
- Developed in 1920's by Arthur Scherbius based patents starting in 1918. There were multiple versions of the Enigma, and other countries had similar machines.
- Based on **automating a polyalphabetic** cipher disk times nine (patch board, rotors, reflector).
- Requires distribution of keywords – WW II Germans issued code books for 30 days, longer for U-boats
- Thought to impregnable based on the large number of set-up configurations.



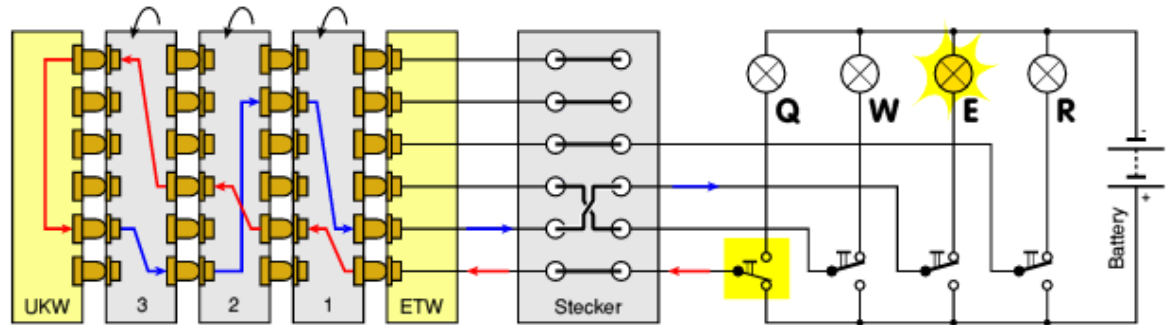
Picture from Wikipedia

German Enigma Machine

How it works: <http://www.cryptomuseum.com/crypto/enigma/working.htm>



Rotor



Simplified circuit diagram of a 3-wheel Service Enigma

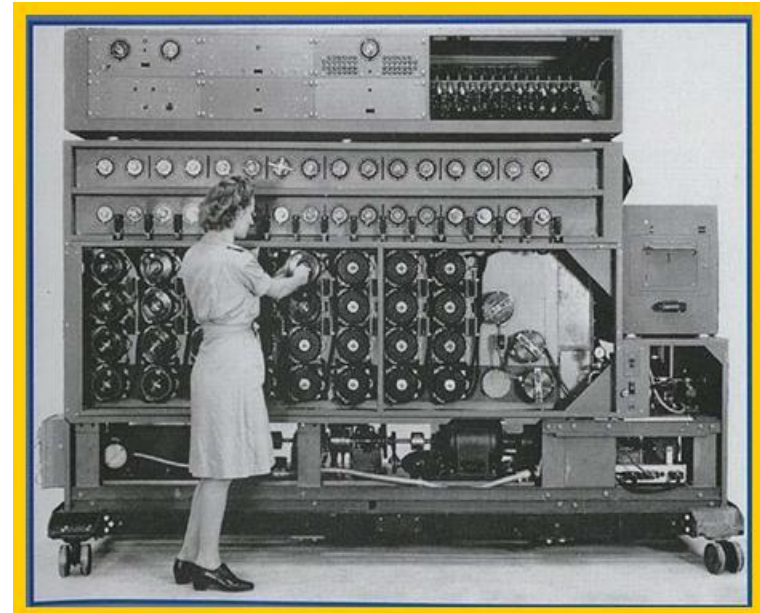
Reflector

Plug board

- After a letter was coded the first rotor would index $1/26^{\text{th}}$ of a rev, and after 1 rev the second rotor would index, etc. The reflector was stationary.
- **How it works video:** <http://www.youtube.com/watch?v=eIYw4Ve4F-I>
- Computer simulators: <http://enigmaco.de/enigma/enigma.html> and <http://users.telenet.be/d.rijmenants/index.htm>

Cracking the Enigma

Cracking Enigma is a spy story better than fiction: In the 1930's a jealous bother leaked documents to the French, they give info to the Poles, Polish mathematician **Marian Rejewski** figured out how to crack it and built a machine (bombe) to do so, on the eve of a German invasion the poles informed the French and English and gave them two enigmas and blueprints for the bombe. In the mean time the Germans added more complexity to the machine that was in use at the start of WW II. Bletchley Park had 9000 workers including **Alan Turing** and many other scientists, mathematicians and linguists from Oxford and Cambridge.



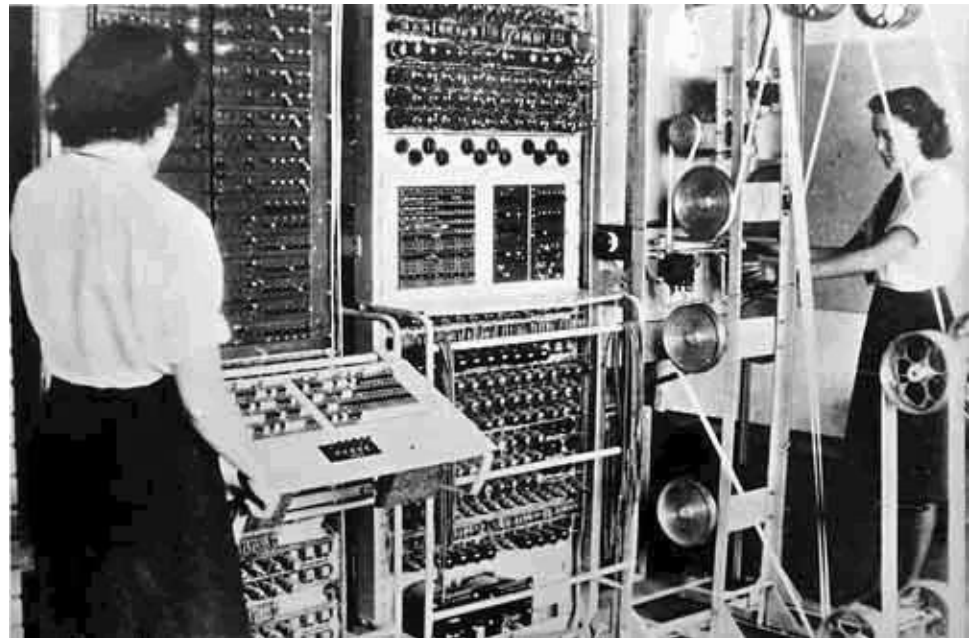
More on Breaking Enigma

- See http://en.wikipedia.org/wiki/Cryptanalysis_of_the_Enigma
- The British bombe (10 min.)
<http://www.youtube.com/watch?v=Hb44bGY2KdU>
- A good overview from the Polish effort to Colossus (23min.)
<http://www.youtube.com/watch?v=eoK4i0SU3DA&feature=related>

Transition to electronic computers at the end of WWII

From Wikipedia: Colossus was the world's first electronic, digital, programmable computer. Colossus and its successors were used by British codebreakers to help read encrypted German messages during World War II. They used thermionic valves (vacuum tubes) to perform the calculations.

The prototype, Colossus Mark 1, was shown to be working in December 1943 and was operational at Bletchley Park by February 1944. An improved Colossus Mark 2 first worked on 1 June 1944, just in time for the Normandy Landings. Ten Colossus computers were in use by the end of the war.





The Adventure of the Dancing Men