A Brief History of Cryptography

Problems from the History of Mathematics

Lecture 23 — April 27, 2018

Brown University

Early Substitution Ciphers

The earliest known ciphers are simple cases of what are now called monoalphabetic substitution ciphers. These include

1. The Atbash Cipher, designed for use with the Hebrew alphabet but extended to other ordered alphabets. One replaces each instance of the first letter with the last, etc:

A	B	C	D	E	F	G	···	T	U	V	W	X	Y	Z
Z	Y	X	W	V	U	T	···	G	F	E	D	C	B	A

2. The Caesar Cipher, named after Julius Caesar, encodes an alphabet by shifting each letter forward a fixed number[1] of places:

A	B	C	D	E	F	G	···	T	U	V	W	X	Y	Z
D	E	F	G	H	I	J	···	W	X	Y	Z	A	C	C

[1] Classically, 3.

General Monoalphabetic Substitution Ciphers

In general, monoalphabetic substitution cipher are keyed by a choice of permutation of the letters in the alphabet. This may seem secure, since there are

$$26! \approx 10^{61}$$

permutations of {A,...,Z} to test in a brute force attack.

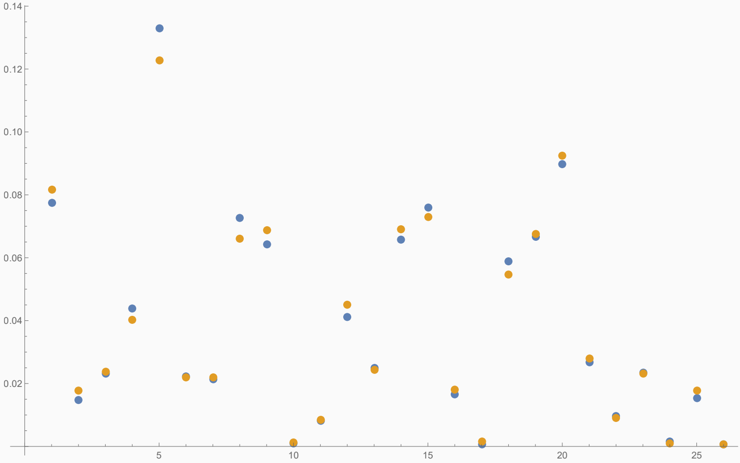
This assumption was proven false by Al-Kindi (c. 800), who pioneered the use of frequency analysis in cryptanalysis.

Monoalphabetic ciphers are now longer used in serious cryptography, but they endure as popular puzzles and are a mainstay of newspapers.[2]

[2] Eg. Cryptoquip.

English Letter Frequencies

The following figure plots the relative frequencies of the letters A-Z in *Madame Bovary* (blue) and *Moby Dick* (orange):



Polyalphabetic Ciphers

The next iteration of ciphers are now known as polyalphabetic ciphers. In these encryption schemes, two letters which agree in the plaintext may encrypt as two different letters.

The most popular polyalphabetic cipher was the Vigenère cipher.[3] In this cipher, an encryption key (re: word) is chosen that determines a periodic sequence of Caesar shifts to apply.

For example, to encrypt

```
PROBLEMS FROM THE HISTORY OF MATHEMATICS
```

with the key `VIGENERE`, we Caesar shift P by V=22 letters to get L; shift R by I=9 letters to get A; etc. We obtain

```
LAVGZJEXBAVRHMWMEBATFDGKIJAMSRSYELZ
```

[3] This is a misattribution. The Vigenère cipher was invented by Giovan Battista Bellosa in 1553.

Cryptanalysis of Polyalphabetic Ciphers

The Vigenère cipher remained secure for three centuries, until Friedrich Kasiski published an attack now known as Kasiski examination.[4]

To apply this technique, we first find all repeated strings in the ciphertext.

```
PROBLEMSFROMTHEHISTORYOFMATHEMATICS
LAVGZJEXBAVRHMWMEBATFDGKIJAMSRSYELZ
 AV     BAV      BA
```

While the two BA arise from different plaintext pairs (FR and ST), the two AV come from the same digraph: RO. The distances between these repetitions are 9 and 8, respectively. If neither is coincidence, we expect the length of the key to be a divisor of 8 or 9. (It was VIGENERE.)

In a longer text we expect many such repetitions. Given such data, we make an educated guess on the key length, perform frequency analysis in congruence classes (mod the key length), and crack the code.

[4] A similar method is described in unpublished work of Babbage from 1846.

Enigma

Classically, innovations in cryptography and cryptanalysis have occurred in times of war.[5] WWII brought along another revolution, marked by the use of rotor machines.

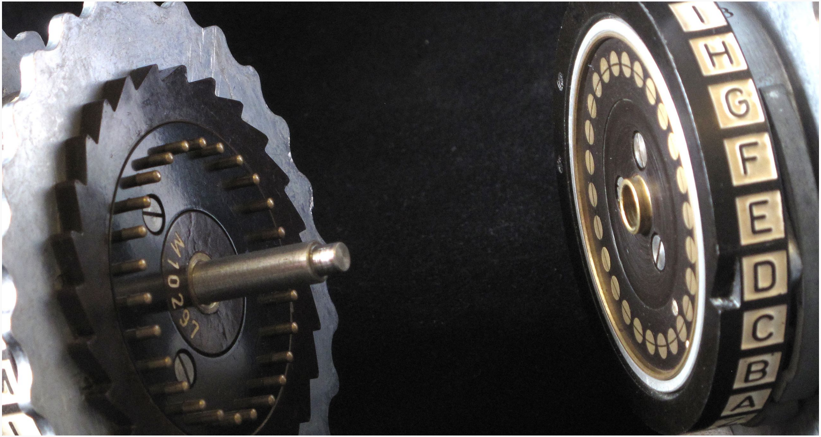
The first rotor machines were invented around 1915 and were defined by the use of rotors to change a monoalphabet cipher with every key press.

Unlike the Vigenère cipher (which also changes the cipher with each new letter), it took tens of thousands of keystrokes to loop back to a previous permutation.

The most famous of these machines, known as Enigma, was designed by Arthur Scherbius from 1918 to 1923.

[5] Babbage's work on polyalphabetical ciphers took place during the Crimean war, eg.

Rotors of an Enigma Machine



Note the electrical contacts and the ratcheted stepping mechanism.

Cryptanalysis of Enigma

The German military adopted the Enigma machine in 1926. In 1932, the Polish mathematician Marian Rejewski reverse engineered Enigma sight-unseen and began to develop attacks against it.

Each message sent with Enigma began with a 3-letter code prescribing the rotor positions used for the rest of the message. This was given two times, but instead of doubling the ciphered trigram the Germans would encrypt it twice.

By comparing the first three characters to characters 4-6, Rejewski could study the cycles in the permutations of the initial positions of the rotors.

This was helpful because Enigma machines included steckers, plugs which complicated Enigma by swapping the circuits of pairs of letters. Choice of steckers does not affect cycle lengths.

The Cryptological Bombe

Rejewski's techniques abbreviated the work of finding daily keys, but some amount of brute force work was still required. Rejewski designed a machine in late 1938 to automate this process, and from December 1938 to mid 1939, the Poles could infer daily settings within two hours.

Following the German invasion of Poland in fall 1939, cryptanalysis of Enigma continued at Bletchley Park in England.[6] This resulted in the creation of Colossus, the first semi-programmable electronic computer.

Later in the war (1942-1943), the United States took over as the leader in bombe manufacturing. Runtime decreased by a factor of 30-40, which compensated for an increase in the number of Enigma rotors.

[6] Germans fixed the Polish exploit in 1940 but work continued with the study of known plaintext, known as cribbing.

Mathematical Cryptography

In 1945, Claude Shannon wrote a paper called *A Mathematical Theory of Cryptography*, which applied information theory[7] to cryptography. Shannon identified the key elements of cryptography, secrecy and authenticity.

Shannon also proved that the one-time pad was had enough information entropy to be cryptographically unbreakable.

[7] Shannon's real claim to fame.

Diffie–Hellman Key Exchange

In 1976, Martin Hellman and Whitfield Diffie published *New Directions in Cryptography*, which described the first (published) private key exchange.

Problem:

How can two people produce a shared secret key over public channels?

Their solution uses modular arithmetic. Suppose that Alice and Bob wish to create a shared secret key.

1. Alice and Bob declare a prime p and a primitive root $g \bmod p$.
2. Alice and Bob choose secret numbers a and b, then exchange $g^a \bmod p$ and $g^b \bmod p$.
3. Alice computes $(g^b)^a \bmod p$ and Bob computes $(g^a)^b \bmod p$. This is the shared secret key.

Diffie–Hellman key exchange relies on the difficulty of the Discrete Log Problem: given p, g, and $g^a \bmod p$, it is hard to recover a.

Questions?