A history of cryptography and cryptanalysis

Christopher Colahan
Simpson College

March 16, 2017

Contents

List of Figures **2**

1 Antiquated Cryptography **3**
1.1 Transposition Ciphers . . . 3
1.2 Monoalphabetic Substitution Ciphers . . . 3
1.2.1 Shift Ciphers . . . 3
1.2.2 Homophonic Substitution Ciphers . . . 3
1.3 Polyalphabetic Substitution Ciphers . . . 4
1.3.1 Vigenère Cipher . . . 4
1.3.2 One Time Pad . . . 5
1.3.3 The Enigma Machine . . . 5
1.4 Cryptanalysis on Substiution Ciphers . . . 5
1.5 Frequency Analysis . . . 5
1.5.1 Frequency Analysis on Monoalphabetic Substitution Ciphers . . . 5
1.5.2 Frequency Analysis on Polyalphabetic Substitution Ciphers . . . 5

2 Modern Cryptography **6**
2.1 One Way Hashing . . . 6
2.1.1 SNEFRU . . . 6
2.1.2 MD5 . . . 6
2.1.3 SHA Family . . . 7
2.2 Cryptanalysis on hashing algorithms . . . 7
2.3 Private Key Cryptography . . . 7
2.4 Public Key Cryptography . . . 7
2.4.1 Certificate Authorities . . . 8

3 Attacks **8**
3.1 Differential Cryptanalysis . . . 8

4 MISC **8**

References **9**

List of Figures

1 Vigenère Square . . . 4
2 Frequency of Characters in English Text . . . 6

All ciphers can be defined as two functions, one called the enciphering function that takes a plaintext as input and outputs a ciphertext, and one called a deciphering function that takes a ciphertext as input and outputs the plaintext.

For convenience, some common notation is used:

p is the plaintext.
c is the ciphertext.
k is the secret key.
$E(p)$ is the enciphering function.
$D(c)$ is the deciphering function.

1 Antiquated Cryptography

1.1 Transposition Ciphers

A transposition cipher is a permutation of the plaintext.

1.2 Monoalphabetic Substitution Ciphers

A monoalphabetic substitution cipher uses a mapping from one alphabet to another.

For convenience, we define a mapping M: $\{A, B, ..., Z\} \to \mathbb{Z}_{26}$ where $A \mapsto 0$, $B \mapsto 1$, ... $Z \mapsto 25$.

1.2.1 Shift Ciphers

Shift ciphers work by shifting the symbols in the plaintext by an amount. For example, if we are using the English alphabet, then there are $n = 26$ possible symbols. We could then choose some k, $0 < k < n$ for our key. In our notation, this would look like

$$E(p_i, k) = M^{-1}(M(p_i) + k \pmod{n}).$$

(possibly cite abs alg textbook here?) To get the deciphering function, we shift backwards:

$$D(c_i, k) = M^{-1}(M(c_i) - k \pmod{n}).$$

1.2.2 Homophonic Substitution Ciphers

A homophonic substitution cipher is a substitution cipher that maps each symbol to one of more symbols in order to prevent frequency analysis from being used.

For example, suppose we have 100 symbols $S = \{s_1, s_2, ...s_{100}\}$. The letter e would map to approximately 12 or those symbols, but the letter a would only map to about 8 of those symbols.

Before enciphering, each letter is replaced at random with one of the symbols it maps to. This means that each symbol in the ciphertext only appears with a frequency of about 1%.

1.3 Polyalphabetic Substitution Ciphers

A polyalphabetic substitution cipher uses multiple monoalphabetic substitution ciphers to generate more possibilities for the ciphertext.

1.3.1 Vigenère Cipher

The vigenère cipher uses 26 alphabets to encrypt plaintext. A key is also used that consists of a string of symbols. Given a plaintext symbol p_i and a key symbol k_j, the ciphertext symbol c_i is the character in the i column and j row. Figure 1 shows the square used for encrypting and decrypting using the Vigenère cipher.

Figure 1: Vigenère 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
A	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	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	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	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	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
F	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
G	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
H	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
I	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
J	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
K	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
L	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
M	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
N	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
O	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
P	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
Q	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
R	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
S	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
T	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
U	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
V	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
W	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
X	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
Y	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
Z	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

The Vigenère Cipher can be mapped to a function easily by inspecting the table. Notice that every character in the A row is the same as the letter in that column. Additionally, every letter in the B row is offset one from the letter in that column, and the C row is offset by 2, and so on.

By observation, we can see that $M(c_i) = M(p_i) + M(k_j)$. Then the encryption and decryption functions for the Vigenère Cipher can be defined as

$$E(p_i, k_j) = M^{-1}(M(p_i) + M(k_j) \pmod{26}) = c_i.$$
$$D(c_i, k_j) = M^{-1}(M(c_i) - M(k_j) \pmod{26}) = p_i.$$

If we know the length of the key is n, then the function could also defined as

$$E(p_i, k_i) = M^{-1}(M(p_i) + M(k_{i \pmod n}) \pmod{26})$$
$$D(c_i, k_i) = M^{-1}(M(c_i) - M(k_{i \pmod n}) \pmod{26})$$

1.3.2 One Time Pad

The one time pad is a special Vigenère cipher where

- the key is the same length as the plaintext,
- the key is random, and
- the same key is not used to encrypt two different plain texts.

There is no statistical analysis that can be applied to the ciphertext [1, pg. 393]. Instead of using a Vigenère cipher, modern implementations use the binary XOR operation to combine ciphertext and plaintext since enciphering and deciphering are the same operation and thus much simper.

1.3.3 The Enigma Machine

The Enigma machine was a German encryption/decryption machine used extensively during world war 2. [insert references].

The Enigma machine consisted of a keyboard, a display, a plugboard, multiple scrambling rotors, and a reflector. When a charcter is entered via the keyboard, an electrical signal is sent first through the plugboard. The plugboard allows pairs of letters to be swapped. Next, the signal goes through the scramblers. Each scrambler acts as a substitution cipher. The reflector then reflects the signal back through the scramblers in the reverse order. Finally, the signal is shown on the corresponding letter on the display. [need citation]

[insert details about breaking enigma]

1.4 Cryptanalysis on Substiution Ciphers

1.5 Frequency Analysis

1.5.1 Frequency Analysis on Monoalphabetic Substitution Ciphers

Shift ciphers are easily broken by frequency analysis. Figure 2 the letter frequency from a sample of English text. If a sufficient sample of cipher-text is acquired, The frequency of letters should be a shifted version of Figure 2.

1.5.2 Frequency Analysis on Polyalphabetic Substitution Ciphers

Breaking a polyalphabetic cipher is more difficut than a monoalphabetic cipher. For breaking a Vigenère cipher, the key length must first be found. To find the key length of a Vigenère cipher, all lengths of keys are checked. If the correct length n is chosen, every n^{th} letter taken together will form a frequency distribution similar to that of a monoalphabetic cipher.

Figure 2: Frequency of Characters in English Text

Letter	Percentage
a	8.2
b	1.5
c	2.8
d	4.3
e	12.7
f	2.2
g	2.0
h	6.1
i	7.0

Letter	Percentage
j	0.2
k	0.8
l	4.0
m	2.4
n	6.7
o	7.5
p	1.9
q	0.1
r	6.0

Letter	Percentage
s	6.3
t	9.1
u	2.8
v	1.0
w	2.4
x	0.2
y	2.0
z	0.1

[3, pg. 19]

This works because every if the key is length n, then every n^{th} character in the key string is the same, thus every n^{th} character in the ciphertext is encrypted using the same alphabet. Once the key length is found, frequency analysis can be used against every n letters in the ciphertext, then every $n-1$ letters, and so on for the length of the key. This process requires the key to be much smaller than the plaintext to be used successfully, so the one-time pad cannot be attacked using this method.

2 Modern Cryptography

2.1 One Way Hashing

One way hashing is a technique commonly used to store passwords. The idea is to take an input set of plaintext P and map it to an output hash set C using the function $H(p) = c$. There should also be no $H^{-1}(c) = p$ (otherwise it would not be one way). To break an ideal one way hash algorithm, the fastest way should be using brute force.

2.1.1 SNEFRU

2.1.2 MD5

MD5, or Message Digest 5, is the fifth improvement to the Message Digest family of hash algorithms designed by Ron Rivest.

The algorithm works as follows. First pad the message with one 1 bit and the rest 0 bits such that $|message| \equiv 448 (\text{mod } 512)$. Then append the 64 bit length of the whole message

to the end. Now $|message| \equiv 0 (\text{mod } 512)$ For each 512 bit block, Let

$$\begin{aligned} m &= \text{input message block (512 bit)} \\ a &= 0x01234567 \\ b &= 0x89ABCDEF \\ c &= 0xFEDCBA98 \\ d &= 0x76543210 \end{aligned}$$

$$\begin{aligned} F(X,Y,Z) &= (X \wedge Y) \vee ((\neg X) \wedge Z) \\ G(X,Y,Z) &= (X \wedge Z) \vee (Y \wedge (\neg Z)) \end{aligned}$$

[2, p. 436]. MD5 has been broken (need citation).

2.1.3 SHA Family

SHA is the Secure Hash Algorithm standard created by NIST and the NSA [2].

2.2 Cryptanalysis on hashing algorithms

Brute force will always work on a hashing algorithm, but it is often infeasible to complete in any meaningful time. Another attack that may be practical is the birthday attack. The *birthday attack* finds 2 input messages m and m' such that $H(m) = H(m')$ [2].

2.3 Private Key Cryptography

In private key cryptography, both users Alice and Bob who wish to communicate securely must have each others secret keys.

2.4 Public Key Cryptography

In public key cryptography each user has two keys, a private key k_{pri} and a public key k_{pub}. k_{pub} is used for encrypting messages, while k_{pri} is used for decrypting messages. Unfortunately, every user must have a list of the public keys for all users they wish to communicate with.

When a user Alice wants to send a message to another user Bob, Alice encrypts the message with Bob's public key. Since Bob is the only one with his private key, he is the only one who can decrypt the message, thus providing secure communication if the algorithm is encryption and decryption algorithm is secure.

Typically, a public key algorithm such as RSA is slow and thus not very feasible fo the real time applications in use today. However, public key cryptography algorithms are commonly used to exchange a key for use with private key algorithms such as AES.

2.4.1 Certificate Authorities

A Certificate Authority (CA) solves the problem of keeping track of keys. Instead of having a key for every other person, the CA keeps them all and each user just has the key of the CA. When the user Alice wants to communicate with Bob, Alice asks the CA for Bob's key via a secure channel. Once Alice gets Bob's key, Alice then can communicate securely with Bob without involving the CA again.

3 Attacks

There are sevaral types of cryptanalytic attacks that can be performed on a cryptographic system [4]:

- *ciphertext only attack*: The cryptanalyst knows only the cipher text
- *known plaintext attack*: The cryptanalyst possesses a substantial quantity of corresponding plaintext and ciphertext
- *chosen plaintext attack*: The cryptanalyst can submit an unlimited number of plaintext messages of their choosing and examine the resulting ciphertext

Kerckhoffs' principle states that a system should be secure when the attacker knows all aspects of the systems except the key. [need citation ?]

3.1 Differential Cryptanalysis

[insert things]

4 MISC

Not sure where these should go.

Possibly put this under polyalphabetic ciphers.

A simple way to expand increase the difficulty of breaking a substitution cipher is to use character pairs, or even more letters. For example, say we define $E(ab) = xg$ and $E(ac) = kd$. If this were a monoalphabetic cipher, then a would need to amp to the same character in both cases. Frequency analysis can still be used, but additional effort is needed since each pair can map to 26^2 pairs. [need citation]. The can be expanded for even larger blocks. For the number of characters in a block, there are 26^n possible mappings.

References

[1] Roberto Tamassia Michael T. Goodrich. *Introduction to Computer Security.*

[2] Bruce Schneier. *Applied Cryptography.*

[3] Simon Singh. *The Code Book.*

[4] Matrin E. Hellman Whitfield Diffie. New directions in cryptography.