Lecture 2: Classical Encryption Techniques

Lecture Notes on "Introduction to Computer Security"

by Avi Kak (kak@purdue.edu)

January 12, 2007

Two Goals:

- To introduce the rudiments of encryption vocabulary.
- To trace the history of some early approaches to cryptography and to show through this history a common failing of humans to get carried away by technological and scientific hubris.

BASIC VOCABULARY OF CLASSICAL ENCRYPTION

plaintext: This is what you want to encrypt

ciphertext: The encrypted output

enciphering or encryption: The process by which plaintext is converted into ciphertext

encryption algorithm: The sequence of data processing steps that go into transforming plaintext into ciphertext. Various parameters used by an encryption algorithm are derived from a secret key.

In classical cryptography for commercial and other civilian applications, the encryption algorithm is made public.

secret key: A secret key is used to set some or all of the various parameters used by the encryption algorithm. **The important thing to note is that the same secret key is used for encryption and decryption in classical cryptography.** It is for this reason that classical cryptography is also referred to as **symmetric key cryptography**.

deciphering or decryption: Recovering plaintext from ciphertext

decryption algorithm: The sequence of data processing steps that go into transforming ciphertext back into plaintext. Various parameters used by a decryption algorithm are derived from the same secret key that was used in the encryption algorithm.

In classical cryptography for commercial and other civilian applications, the decryption algorithm is made public.

cryptography: The many schemes available today for encryption and decryption

cryptographic system: Any single scheme for encryption

cipher: A cipher means the same thing as a "cryptographic system"

block cipher: A block cipher processes a block of input data at a time and produces an ciphertext block of the same size.

stream cipher: A stream cipher encrypts data on the fly, usually one byte at at time.

cryptanalysis: Means "breaking the code". Cryptanalysis relies on a knowledge of the encryption algorithm (that for civilian applications should be in the public domain) and some knowledge of the possible structure of the plaintext (such as the structure of a typical inter-bank financial transaction) for a partial or full reconstruction of the plaintext from ciphertext. Additionally, the goal is to also infer the key for decryption of future messages.

The precise methods used for cryptanalysis depend on whether the "attacker" has just a piece of ciphertext, or pairs of plaintext and ciphertext, how much structure is possessed by the plaintext, and how much of that structure is known to the attacker.

All forms of cryptanalysis for classical encryption exploit the fact that some aspect of the structure of plaintext may survive in the ciphertext.

brute-force attack: When encryption and decryption algorithms are publicly available, a brute-force attack means trying every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained.

key space: The total number of all possible keys that can be used in a cryptographic system. For example, **DES** uses a 56-bit key. So the key space is of size 2^{56}, which is approximately the same as 7.2×10^{16}.

cryptology: Cryptography and cryptanalysis together constitute the area of cryptology

Building Blocks of Classical Encryption Techniques

- Two building blocks of all classical encryption techniques are **substitution** and **transposition**.

- Substitution means replacing an element of the plaintext with an element of ciphertext.

- Transposition means rearranging the order of appearance of the elements of the plaintext.

- Transposition is also referred to as permutation.

Caesar Cipher

- This is the earliest known example of a substitution cipher.

- Each character of a message is replaced by a character three position down in the alphabet.

```
plaintext:   are you ready

ciphertext:  DUH BRX UHDGB
```

- If we represent each letter of the alphabet by an integer that corresponds to its position in the alphabet, the formula for replacing each character 'p' of the plaintext with a character 'C' of the ciphertext can be expressed as

```
C  =  E( 3, p )  =  (p + 3) mod 26
```

- A more general version of this cipher that allows for any degree of shift would be expressed by

```
C  =  E( k, p )  =  (p + k) mod 26
```

- The formula for decryption would be

```
p  =  D( k, C )  =  (C - k) mod 26
```

- In these formulas, 'k' would be the secret key. The symbols 'E' and 'D' represent encryption and decryption.

The Swahili angle ...

- A simple substitution cipher obviously looks much too simple, but that is the case only if you have some idea regarding the nature of the plaintext.

- What if the "plaintext" could be considered to be a binary stream of data and a substitution cipher replaced every consecutive 6 bits with one of 64 possible cipher characters? *In fact, this is referred to as Base64 encoding for sending email multimedia attachments.*

- If you did not know anything about the underlying plaintext and it was encrypted by a Base64 sort of algorithm, it might not be as trivial a cryptographic system as the substitution cipher shown on the previous page. But, of course, if the word ever got out that your plaintext was in Swahili, you'd be hosed.

Monoalphabetic Ciphers

- In a monoalphabetic cipher, our substitution characters are a random permutation of the 26 letters of the alphabet:

```
plaintext letters:        a  b  c  d  e  f .....

substitution letters:     t  h  i  j  a  b .....
```

- The key now is the sequence of substitution letters. In other words, the key in this case is the actual random permutation of the alphabet used.

- Note that there are `26!` permutations of the alphabet. That is a number larger than 4×10^{26}.

A Very Large Key Space But ....

- That gives us a huge key space (meaning the total number of all possible keys that would need to be guessed in a brute-force attack). This key space is 10 orders of magnitude larger than the size of the key space for DES, the now somewhat outdated (but still widely used) NIST standard.

- Obviously, this would rule out a brute-force attack. (Even if each key took only a nanosecond to try, it would still take zillions of years to try out even half the keys.)

- So this would seem to be the answer to our prayers for an unbreakable code for symmetric encryption.

- But it is not!

- Why?

The All-Fearsome Statistical Attack

- If you know the nature of plaintext, any substitution cipher, regardless of the size of the key space, can be broken easily with a statistical attack.

- When the plaintext is plain English, a simple form of statistical attack consists measuring the frequency distribution for single characters, for pairs of characters, for triples of characters, etc., and comparing those with similar statistics for English.

- The figure on the next page shows the relative frequency of of the letters in a sample of English text. Obviously, by comparing this distribution with a histogram for the characters in a piece of ciphertext, you may be able to establish the true identities of the ciphertext characters.

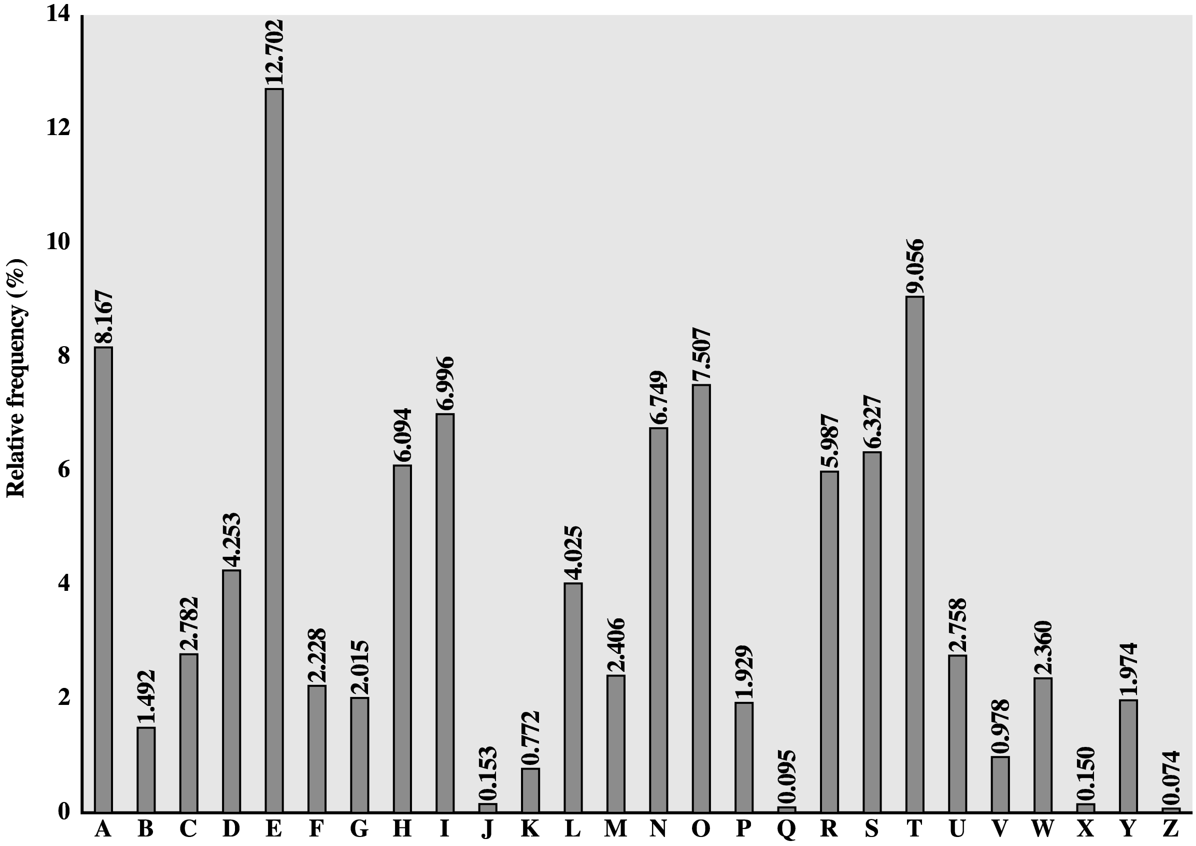

Figure 2.5 Relative Frequency of Letters in English Text

This figure is from Chapter 2 of Stallings: “Cryptography and Network Security”, Fourth Edition

Comparing the Statistics for Digrams and Trigrams

- Equally powerful statistical inferences can be made by comparing the relative frequencies for pairs and triples of characters in the ciphertext and the language believed to be used for the plaintext.

- Pairs of adjacent characters are referred to as **digrams**, and triples of characters as **trigrams**.

- To illustrate the usefulness of digram and trigram occurrences, the most frequently occurring trigram in English text is “the” and the most frequently occurring digram “th”.

Multiple-character Encryption to Mask Plaintext Structure

- One character at a time substitution obviously leaves too much of the plaintext structure in ciphertext.

- So how about destroying some of that structure by mapping multiple characters at a time to ciphertext characters?

- The best known approach that carries out multiple-character substitution is known as **Playfair cipher**.

Constructing the Matrix for Pairwise Substitutions in Playfair Cipher

In Playfair cipher, you first choose an encryption key. You then enter the letters of the key in the cells of a 5 × 5 matrix in a right to left fashion starting with the first cell at the top-left corner. You fill the rest of the cells of the matrix with the remaining letters in alphabetic order. The letters I and J are assigned the same cell. In the following example, the key is "**smythework**":

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

Substitution Rules for Pairs of Characters in Playfair Cipher

1. Two plaintext letters that fall in the same row of the 5×5 matrix are replaced by letters to the right of each in the row. The “rightness” property is to be interpreted circularly in each row, meaning that the first entry in each row is to the right of the last entry. Therefore, the pair of letters “bf” in plaintext will get replaced by “CA” in ciphertext.

2. Two plaintext letters that fall in the same column are replaced by the letters just below them in the column. The “belowness” property is to be considered circular, in the sense that the topmost entry in a column is below the bottom-most entry. Therefore, the pair “ol” of plaintext will get replaced by “CV” in ciphertext.

3. Otherwise, for each plaintext letter in a pair, replace it with the letter that is in the same row but in the column of the other letter. Consider the pair “gf” of the plaintext. We have ’g’ in the fourth row and the first column; and ’f’ in the third row and the fifth column. So we replace ’g’ by the letter in the same row as ’g’ but in the column that contains ’f’. This given us ’P’ as a replacement for ’g’. And we replace ’f’ by the letter in the same row as ’f’ but in the column that contains ’g’. That gives us ’A’ as replacement for ’f’. Therefore, ’gf’ gets replaced by ’PA’.

Dealing with Duplicates Letters in a Key and Repeating Letters in Plaintext

- You must drop any duplicates in a key.

- Before the substitution rules are applied, you must insert a chosen “filler” letter (let’s say it is ’x’) between any repeating letters in the plaintext. So a plaintext word such as “hurray” becomes “hurxray”

How Secure is the Playfair Cipher?

- Playfair was thought to be unbreakable for many decades.

- It was used as the encryption system by the British Army in World War 1. It was also used by the U.S. Army and other Allied forces in World War 2.

- But, as it turned out, Playfair was extremely easy to break.

- As expected, the cipher does alter the relative frequencies associated with the individual letters and with digrams and with trigrams, but not sufficiently.

- The figure on the next page shows the single-letter relative frequencies in descending order (and normalized to the relative frequency of the letter 'e') for different ciphers.

  So whereas the individual letters in the output do occupy a much greater range for Playfair, there is still considerable information left in the distribution for good guesses.

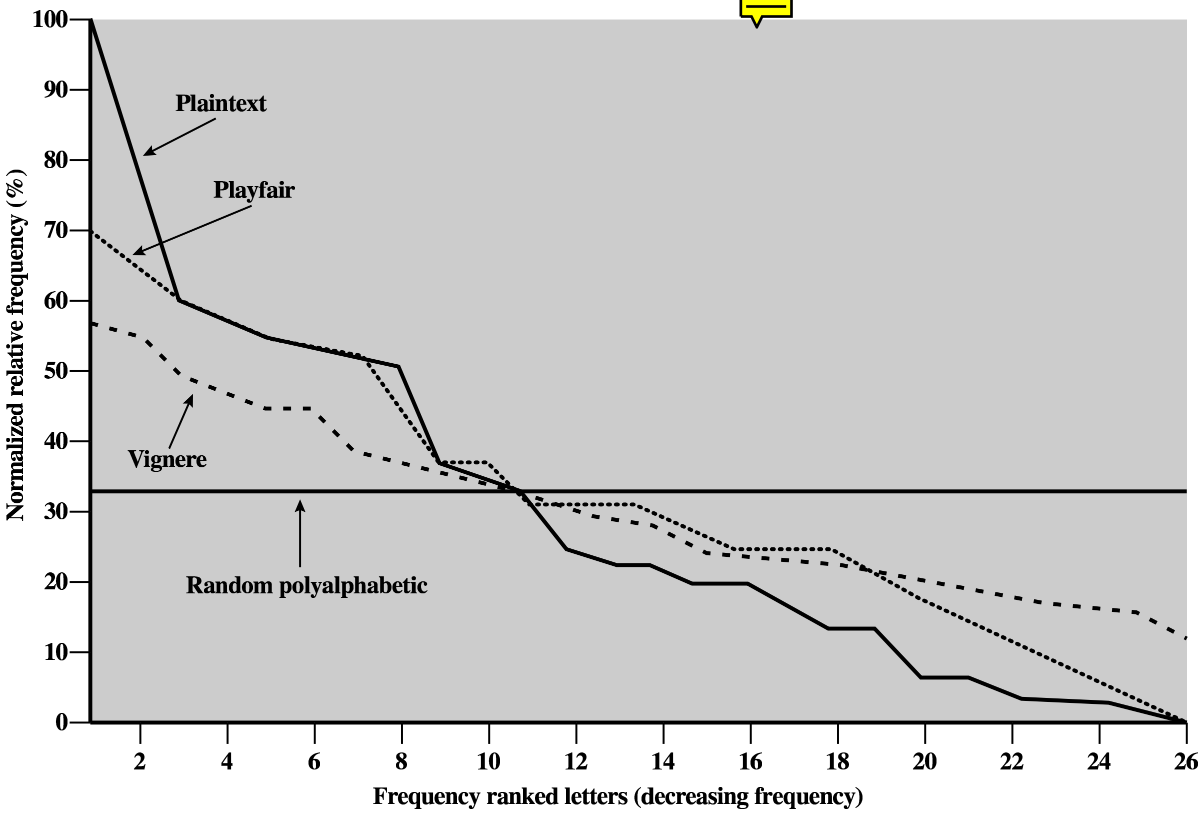

Figure 2.6 Relative Frequency of Occurrence of Letters

This figure is from Chapter 2 of Stallings: “Cryptography and Network Security”, Fourth Edition

Another Multi-Letter Cipher: The Hill Cipher

The Hill cipher takes a very different (more mathematical) approach to multi-letter substitution:

- You assign integers to each letter of alphabet. For the sake of discussion, let's say that you have assigned the integers 0 through 25 to the letter a through z of the plaintext.

- The encryption key, call it **K**, consists of a 3×3 matrix of integers:

$$\mathbf{K} \quad = \quad \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$$

- Now we can transform three letters at a time from plaintext, the letters being represented by the numbers p_1, p_2, and p_3, into three ciphertext letters c_1, c_2, and c_3 in their numerical representations by

$$\begin{aligned} c_1 &= (\ k_{11}p_1 \ + \ k_{12}p_2 \ + \ k_{13}p_3\ )\ mod\ 26 \\ c_2 &= (\ k_{21}p_1 \ + \ k_{22}p_2 \ + \ k_{23}p_3\ )\ mod\ 26 \\ c_3 &= (\ k_{31}p_1 \ + \ k_{32}p_2 \ + \ k_{33}p_3\ )\ mod\ 26 \end{aligned}$$

- The above set of linear equations can be written more compactly in the following vector-matrix form:

$$\vec{\mathbf{C}} \ = \ [\mathbf{K}]\, \vec{\mathbf{P}}\ mod\ 26$$

- Obviously, the decryption would require the inverse of **K** matrix.

$$\vec{\mathbf{P}} \ = \ \left[\mathbf{K}^{-1}\right] \vec{\mathbf{C}}\ mod\ 26$$

This works because

$$\vec{\mathbf{P}} \ = \ \left[\mathbf{K}^{-1}\right] [\mathbf{K}]\, \vec{\mathbf{P}}\ mod\ 26 \ = \ \vec{\mathbf{P}}$$

How Secure is the Hill Cipher?

- It is extremely secure against ciphertext only attacks. That is because the keyspace can be made extremely large by choosing the matrix elements from a large set of integers. (The key space can be made even large by generalizing the technique to larger-sized matrices.)

- But it has zero security when the plaintext–ciphertext pairs are known. The key matrix can be calculated easily from a set of known $\vec{\mathbf{P}}$, $\vec{\mathbf{C}}$ pairs.

Polyalphabetic Ciphers: The Vigenere Cipher

- In a monoalphabetic cipher, the same substitution rule is used for every substitution. In a polyalphabetic cipher, the substitution rule changes continuously from letter to letter according to the elements of the encryption key.

- Let each letter of the encryption key denote a shifted Caesar cipher, the shift corresponding to the key. This is shown on the next page.

- Now a plaintext message may be encrypted as follows

```
key:          abracadabraabracadabraabracadabraab
plaintext:    canyoumeetmeatmidnightihavethegoods
ciphertext:   CBEYQUPEFKMEBK.....................
```

- The Vigenere cipher is an example of a polyalphabetic cipher.

- Since, in general, the encryption key will be shorter than the message to be encrypted, for the Vignere cipher the key is repeated, as illustrated in the above example where the key is the string "abracadabra".

encryption key letter	*plain text letters*				
	a	b	c	d	....
	substitution letters				
a	A	B	C	D	....
b	B	C	D	E	....
c	C	D	E	F	....
d	D	E	F	G	....
e	E	F	G	H	....
.	.	.	.	.	.
.	.	.	.	.	.
z	Z	A	B	C	....

How Secure is the Vigenere Cipher?

- Since there exist in the output multiple ciphertext letters for each plaintext letter, you would expect that the relative frequency distribution would be effectively destroyed. But as can be seen in the plots on page 19, a great deal of the input statistical distribution still shows up in the output. (The plot shown for Vigenere cipher is for an encryption key that is 9 letters long.)

- Obviously, the longer the encryption key, the greater the masking of the structure of the plaintext. The best possible key is as long as the plaintext message and consists of a purely random permutation of the 26 letters of the alphabet. This would yield the ideal plot shown in the figure on page 19 of these notes.

- In general, to break the Vigenere cipher, you first try to estimate the length of the encryption key. This length can be estimated by using the logic that plaintext words separated by multiples of the length of the key will get encoded in the same way.

- If the estimated length of the key is N, then the cipher consists of N monoalphabetic substitution ciphers and the plaintext letters at positions 1, N, 2N, 3N, etc., will get encoded by the same monoalphabetic cipher. This insight can be used to decode each monoalphabetic cipher.

TRANSPOSITION TECHNIQUES

- All of our discussion so far has dealt with substitution ciphers. We have talked about monoalphabetic substitutions, polyalphabetic substitutions, etc.

- We will now talk about a different notion in classical cryptography: permuting the plaintext.

- This is how a pure permutation cipher could work: You write your plaintext message along the rows of a matrix of some size. You generate ciphertext by reading along the columns. The order in which you read the columns is determined by the encryption key:

```
key:              2 5 3 1 6 4

plaintext:        m e e t m e
                  a t m i d n
                  i g h t f o
                  r t h e g o
                  d i e s x y

ciphertext:       ETGTIMDFGXEMHHEMAIRDENOOYTITES
```

- The cipher can be made more secure by performing multiple rounds of such permutations.