CS 555: Cryptography

Historically Classic Ciphers
Shift Cipher

- **Key Space:**
  - [1 .. 25]

- **Encryption given a key K:**
  - each letter in the plaintext P is replaced with the K’th letter following it in the circular list of alphabet symbols (“K-shift to right”)

- **Decryption given K:**
  - “K-shift to left”

---

History: K = 3, Caesar’s 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 |
| 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |

P = **CRYPTOGRAPHYISFUN**  
K = **11**  
C = **NCJAVZRCI2ASJTDQFY**
Shift Cipher: Cryptanalysis

- Can an attacker find $K$?
  - YES: exhaustive search, because key space is too small (25 possible keys)

- Once $K$ is found, it is very easy to decrypt
General Mono-alphabetical Substitution Cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \ldots, Z\}$
- Encryption given a key $\pi$:
  - each letter $X$ in the plaintext $P$ is replaced with $\pi(X)$
- Decryption given a key $\pi$:
  - each letter $Y$ in the cipherext $P$ is replaced with $\pi^{-1}(Y)$

Example:

$$
\begin{array}{cccccccccccccccccc}
\end{array}
$$

BECAUSE $\rightarrow$ AZDBJSZ
Strength of the General Substitution Cipher

- Exhaustive search is impractical
  - key space size is $26! \approx 4 \times 10^{26}$
- Dominated the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then (because of large key space)
- Flawed even though the key space is large
Cryptanalysis of Substitution Ciphers: Frequency Analysis

- Each language has certain features: frequency of letters, or of groups of two or more letters
- Substitution ciphers preserve the language features
- Substitution ciphers are vulnerable to frequency analysis attacks.
Frequency of Letters in English
Other Frequency Features of English

- Vowels, which constitute 40% of plaintext, are often separated by consonants.
- Letter A is often found in the beginning of a word or second from last.
- Letter I is often third from the end of a word.
- Letter Q is followed only by U
- And more …
The number of different ciphertext characters or combinations are counted to determine the frequency of occurrence.

The ciphertext is examined for patterns, repeated series, and common combinations.

Replace ciphertext characters with possible plaintext equivalents using known language characteristics.
Towards the Polyalphabetic Substitution Ciphers

- Main weaknesses of monoalphabetic substitution ciphers
  - each letter in the ciphertext corresponds to only one letter in the plaintext letter

- Idea for a stronger cipher (1460s by Alberti)
  - use more than one cipher alphabet, and switch between them when encrypting different letters

- Developed into a practical cipher by Vigenère (published in 1586)
The Vigenère Cipher

**Definition:**
Key of length m:  \( K = (k_1, k_2, \ldots, k_m) \). To encrypt a long text, use \( K \) on leftmost m symbols of text, then on next m, etc.

**Encryption:**
\[
e_k(p_1, p_2 \ldots p_m) = (p_1+k_1, p_2+k_2 \ldots p_m+k_m) \pmod{26}
\]

**Decryption:**
\[
d_k(c_1, c_2 \ldots c_m) = (c_1-k_1, c_2-k_2 \ldots c_m-k_m) \pmod{26}
\]

**Example with m=4 and n=12:**

Key:         \( \text{L U C K} \)
Plaintext:   \( \text{C R Y P T O G R A P H Y} \)
             \( \text{L U C K L U C K L U C K} \)
Ciphertext: \( \text{N L A Z E I I B L J J I} \)
Security of Vigenere Cipher

- Vigenere masks the frequency with which a character appears in a language: one letter in the ciphertext corresponds to multiple letters in the plaintext. Makes the use of frequency analysis more difficult.
- Any message encrypted by a Vigenere cipher is a collection of as many shift ciphers as there are letters in the key.
Vigenere Cipher: Cryptanalysis (1)

- Find the length $m$ of the key, then
- divide the message into $m$ shift cipher encryptions (each a subsequence whose chars are $m$ positions apart in the message)
- then use frequency analysis to solve the resulting $m$ shift ciphers
Vigenere Cipher: Cryptanalysis (2)

- How to find the length $m$ of the key
- Try various “candidates” for $m$: For each candidate value $t$ count how many times symbols that are $t$ positions apart equal each other
- A “spike” in the count occurs when $t$ is a multiple of $m$ (why?)
One-Time Pad: Vigenere with m=n

Key is chosen randomly
Plaintext $X = (x_1 \ x_2 \ \ldots \ x_n)$
Key $K = (k_1 \ k_2 \ \ldots \ k_n)$
Ciphertext $Y = (y_1 \ y_2 \ \ldots \ y_n)$

$e_k(X) = (x_1 + k_1 \ x_2 + k_2 \ \ldots \ x_n + k_n) \mod s$
$d_k(Y) = (y_1 - k_1 \ y_2 - k_2 \ \ldots \ y_n - k_n) \mod s$

If alphabet size $s$ is 2 (i.e., binary) then both the + and – are the XOR operation