CS 6160 Cryptology Lecture 2: Classical Ciphers and Perfect Secrecy

Maria Francis

August 26, 2021

Caeser Cipher/Shift Cipher

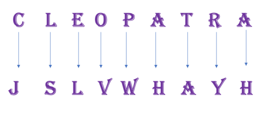
- Named after Julius Ceaser who used it to communicate with his generals.
- Replace each letter with one that is a fixed number of places down the alphabet.

Ceasar cipher

$$\mathcal{M} = \{A, B, \ldots, Z\}^*$$
$$\mathcal{K} = \{0, 1, 2, \ldots, 25\}$$
$$Gen = k, k \in \mathcal{K}$$
$$Enc_k(m_1 m_2 \ldots m_n) = (c_1 c_2 \ldots c_n), \text{ where } c_i = m_i + k \text{ mod } 26$$
$$Dec_k(c_1 c_2 \ldots c_n) = (m_1 m_2 \ldots m_n) \text{ where } m_i = c_i - k \text{ mod } 26.$$

Caeser Cipher/Shift Cipher

- Encrypted messages look scrambled (unless k is known).
- Encrypt with $k = 7$



- Cryptanalysis
  - We just need to try all 26 different values of k see if the resulting plaintext is readable.
  - If the message is relatively long, the scheme is easily broken.

Substitution Cipher

- Choose a permutation π of the alphabet set $\{A, B \ldots, Z\}$ and apply that to all letters in the plaintext.
- Permutation : one-one, onto function from a set to itself
- Brute-force won't work – you have to try $26! \approx 2^{88}$ possible keys.

Substitution Cipher

$$\mathcal{M} = \{A, B, \ldots, Z\}^*$$

$$\mathcal{K} = \text{the set of permutations of } \{A, B \ldots, Z\}$$

$$Gen = \pi, \pi \in \mathcal{K}$$

$$Enc_\pi(m_1 m_2 \ldots m_n) = c_1 c_2 \ldots c_n, \text{ where } c_i = \pi(m_i)$$

$$Dec_\pi(c_1 c_2 \ldots c_n) = m_1 m_2 \ldots m_n \text{ where } m_i = \pi^{-1}(c_i).$$

Cryptanalysis of Substitution Cipher ?

Different types of attacks

- Passive attack – Ciphertext-only attack : Attack performed with only ciphertexts. Most difficult attack.
- Passive attack – Known-plaintext attack (KPA): Attacker is given the pair (plaintext, ciphertext). Relevant because the attacker may know side information (e.g: headers) that allows him to deduce some plaintexts.
- Active attack – Chosen-plaintext attack (CPA): Attacker obtains (plaintext, ciphertext) where plaintexts are of his choice. e.g: information we encrypt is not guaranteed to come from trusted sources.
- Active attack – Chosen-ciphertext attack (CCA): Attacker requests (plaintext, ciphertext) for arbitrary ciphertexts of his choice. E.g: We cannot always trust the provenance of the ciphertexts we decrypt.

Cryptanalysis of Substitution Cipher

- Chosen plaintext attack - completely insecure!
- Ciphertext only (passive) attack? – Frequency analysis
- E.g: in the ciphertext, if x is the most common letter it is likely that $\pi(e) = x$.

a	0.0804	h	0.0549	o	0.0760	v	0.0099
b	0.0154	i	0.0726	p	0.0200	w	0.0192
c	0.0306	j	0.0016	q	0.0011	x	0.0019
d	0.0399	k	0.0067	r	0.0612	y	0.0173
e	0.1251	l	0.0414	s	0.0654	z	0.0009
f	0.0230	m	0.0253	t	0.0925		
g	0.0196	n	0.0709	u	0.0271		

Probability distributions of 1-grams in English.

Additionally, we need to make use of the frequencies of digrams (two letter seq.) and trigrams (three letter seq.) in the plaintext language. For e.g. frequent three letter words : “and”, “the”.

Vigenère cipher

- So far, all were monoalphabetic ciphers – each symbol in the plaintext is mapped to a unique symbol in the ciphertext based on the secret key.
- Vigenère cipher is a polyalphabetic cipher – same plaintext symbol can be mapped to more than one ciphertext symbols.
- A generalization of the shift cipher where each letter of the plaintext is shifted by different amounts.
- Key is a string $k = k_1 \dots k_n$ with $k_i \in \{0, \dots, 25\}$
- Encryption of $m = m_1 \dots m_l$ under key k is $(m_1 + k_1 \bmod 26)(m_2 + k_2 \bmod 26) \dots (m_n + k_n \bmod 26)(m_{n+1} + k_1 \bmod 26), \dots)$.

Vigenère cipher

$$\mathcal{M} = \{A, B, \dots, Z\}^*$$
$$\mathcal{K} = \{k = (k_1 \dots k_n) : k_i \in \{0, \dots, 25\}\},$$
$$Gen = k, k \in \mathcal{K}$$
$$Enc_k(m_1 m_2 \dots m_l) = c_1 c_2 \dots c_l, \text{ where } c_i = m_i + k_j \bmod 26,$$
$$j = i \bmod n$$
$$Dec_k(c_1 c_2 \dots c_l) = m_1 m_2 \dots m_l \text{ where } m_i = c_i - k_j \bmod 26,$$
$$j = i \bmod n$$

S	E	N	D	R	E	I	N	F	O	R	C	E	M	E	N	T	S
V	I	G	E	N	E	R	E	V	I	G	E	N	E	R	E	V	I
N	M	T	H	E	I	Z	R	A	W	X	G	R	Q	V	R	O	A

Cryptanalysis of Vigenère cipher

- If both the plaintext and the ciphertext are known, it is easy to break the system. Just compute the difference between each letter in the ciphertext and the plaintext.
- And insecure of course with a chosen plaintext attack.
- What about ciphertext only attack?
- The key space is of size 26^n so except for small n a brute force attack is not possible.
- Frequency distribution wont work.
- Charles Babbage and "Kasiski Test" (Both came up independently and Babbage was earlier.)

'Kasiski Test"

First step - determining n

- Determine the keyword length n.
- Any two (or more) identical segments of plaintext will encrypt to the same ciphertext letters whenever the distance is a multiple of n.
- Look for identical segments of the ciphertext.
  - Suppose we have m such identical segments.
  - Record the distance between starting position of two segments say $l_1, l_2, \ldots$
  - Prove : n divides l_1, l_2 and n divides the gcd of $l_1, l_2, \ldots,$ and therefore n is the GCD.

‘Kasiski Test”

Another way to determine n

- Guess for n and divide the ciphertext into n bins - $B_0, B_1, \ldots, B_{n-1}$ by placing the ith ciphertext into $B_{i \bmod n}$.
- If the frequency distribution of the symbols n each bin resembles the expected distribution of a “meaningful" English text, then our guess is most probably correct.

‘Kasiski Test”

Second step - determining the keyword

- Suppose we have got the correct keyword length n and the ciphertext symbols are arranged in bins $B_0, \ldots, B_{n-1}$ as in Strategy II.
- The ciphertext symbols in each bean B_i is the result of applying a “shift cipher" (i.e., a cyclic shift of the corresponding plaintext letters.)
- Use the frequency distribution of ciphertext symbols in B_i to make a guess for the ith letter of the keyword.
- Use partial guesses for the key letters to guess the keyword.

Vernam Cipher – One Time Pad

$$\mathcal{M} = \{0,1\}^*$$

$$\mathcal{K} = \{0,1\}^* \text{ where key length} = \text{message length}$$

$$Gen = k, k \in \mathcal{K}$$

$$Enc_k(m_1 m_2 \dots m_n) = c_1 c_2 \dots c_n, \text{ where } c_i = m_i \oplus k_i$$

$$Dec_k(c_1 c_2 \dots c_n) = m_1 m_2 \dots m_n \text{ where } m_i = c_i \oplus k_i$$

- Vigenère cipher with key length equal to the length of the plaintext.
- Key must be chosen in a completely random way and only used once.
- Perfectly secret but impractical! Key should be as long as message and used only once.

One Time Pad

- Encrypting and Decrypting : just XOR with the secret!

$$Enc_k(m) = c = m \oplus k$$
$$Dec_k(c) = m = c \oplus k$$

- Why is it secure? Every $m \in \mathcal{M}$ and ciphertext $c \in \mathcal{C}$ correspond to a unique key k
- What is perfect secrecy?
  A method is secure iff the odds of the adversary to figure out m are the same whether or not he has seen c.
- How to formalize this notion?

Perfectly Secret Encryption

Definition

Let $m \in \mathcal{M}$ be a random message and $c \in \mathcal{C}$ be the ciphertext of m. The encryption scheme is said to be perfectly secure if for an adversary $Pr[M = m|C = c] = Pr[M = m]$.

One Time Pad is Perfectly Secure

Proof: To show that $Pr[M = m|C = c] = Pr[M = m]$ for each pair m, c.

$$Pr[(M = m|C = c)] = \frac{Pr[(M = m \cap C = c)]}{Pr[C = c]}$$

by Bayes law,

$$= \frac{Pr[(M = m)] \cdot Pr[(C = c|M = m)]}{Pr[C = c]}$$

by conditional prob. def.,

$$= \frac{Pr[(M = m)] \cdot Pr[(C = c|M = m)]}{\sum_{m' \in \mathcal{M}}(Pr[M = m'] \cdot Pr[C = c|M = m'])}$$

by expanding Pr[C=c] as the sum of all cond. prob.

Proof Contd

- Note that $Pr[C = c|M = m^{'}] = Pr[k = c \oplus m^{'}]$ in OTP.
- Since every $k \in \{0,1\}^n$ is equally likely to be a key $Pr[k = c \oplus m^{'}] = \frac{1}{2^n}$.

$$
\begin{aligned}
Pr[M = m|C = c] &= \frac{Pr[M = m] \cdot \frac{1}{2^n}}{\sum_{m' \in \mathcal{M}} (Pr[M = m']) \cdot \frac{1}{2^n}} \\
&= \frac{Pr[M = m]}{\sum_{m' \in \mathcal{M}} (Pr[M = m'])} \\
&= \frac{Pr[M = m]}{1}
\end{aligned}
$$

Shannon's result

- OTPs are not practical especially because of the key length.
- Can we have a clever way of getting perfect secrecy with shorter keys? Unfortunately the answer is no!

Theorem (Shannon)

For any perfectly secure scheme where Alice and Bob share a key k from space $\mathcal{K}$ and can encrypt any message m from space $\mathcal{M}$, we must have $|\mathcal{K}| \geq |\mathcal{M}|$.

Thus OTP is optimal in this regard. Anybody else claiming that they have discovered an unbreakable cipher with shorter keys are wrong!

Shannon's result - Proof

- For any valid ciphertext c, let A be the number of messages m that could result from the decryption of c under some secret key k.
- Let us estimate A in two ways:
- For a given key $k \in \mathcal{K}$ there can be at most one m since Alice could decrypt c in at most one way for each k.
- Thus $|A| \leq |\mathcal{K}|$.
- Claim : $|A| = |\mathcal{M}|$, i.e. every $m \in \mathcal{M}$ can result in producing c.
- If not for some m, then $Pr[M = m] > 0$ before we saw c, but $Pr[M = m|C = c] = 0$, contradiction to perfect security!
- Thus, $|A| = |\mathcal{M}| \leq |\mathcal{K}|$.

Observations

- Perfect secrecy is w.r.t. computationally unbounded adversary. This is why we assumed Eve to be a PPT.
- Is this true? : Every encryption scheme for which $|\mathcal{K}|$ equals $|\mathcal{M}|$ and for which the key is chosen uniformly from $\mathcal{K}$, is perfectly secret. A: False.
  - Let $\mathcal{M} = \{a, b\}$, $\mathcal{K} = \{k_1, k_2\}$, $\mathcal{C} = \{0, 1\}$.
  - Let $Enc_k(a) = 0$ and $Enc_k(b) = 1$ for $k = k_1, k_2$.
  - Dec algorithm will return a on input ciphertext 0 and b on input ciphertext 1.
  - Clearly, the scheme is correct.

$$Pr[M = a | C = 1] = 0 \neq (1/2) = Pr[M = a]$$

  not perfectly secret!
- *Gen* must choose the key uniformly from the set of all keys but that is not enough! for every message m and ciphertext c there is a unique key mapping m to c

Observations/Exercises

- Ceaser cipher is definitely not secure. What if we encrypt only one letter? i.e., $\mathcal{M} = \mathcal{C} = \{0, \ldots, 25\}$ and not $\{0, 1, \ldots, 25\}^*$? Prove that in such a scenario it is a perfectly secure cipher!
- Consider an encryption scheme (Gen, Enc, Dec) where for any two messages $m, m' \in \mathcal{M}$ the distribution of the ciphertext when m is encrypted is identical to the distribution of the ciphertext when m' is encrypted. i.e.

$$Pr[Enc_K(m) = c] = Pr[Enc_K(m') = c], \forall c \in \mathcal{C} \tag{1}$$

  The encryption scheme is said to have adversarial indistinguishability .
- Q: Show that it is equivalent to saying an encryption scheme is perfectly secret.