SINGAPORE INSTITUTE OF TECHNOLOGY

ICT2213 Applied Cryptography

Topic 1.2: Classical Cryptography (Mono-alphabetic ciphers)

Learning outcomes

- Explain how classical mono-alphabetic ciphers, such as the shift and substitution ciphers, work
- Use statistical attacks to cryptanalyze these ciphers

Shift cipher

Operation

- Shift each plaintext character forward by k places, where k is the key
- If k = 3, a becomes D, b becomes E, etc.
- Characters are wrapped around so, for k = 3, z becomes C, etc.
- To decrypt, we shift backwards by the same amount (the key)

Example

- Assume the key is 4, and the plaintext is **begintheattacknow**
- Note that, we have removed the spaces from the input plaintext
- The resulting ciphertext is **FIKMRXLIEXXEGORSA**
- Observe that w wrapped back to become A
- Typically, we show the plaintext in lowercase and the ciphertext in uppercase

Properties of the shift cipher

Mono-alphabetic cipher

- The shift cipher is a mono-alphabetic cipher
- That is, each character of the plaintext is replaced with a corresponding character of the ciphertext
- One-to-one mapping

Mathematical notation

- The shift-with-wrap-around operation translates nicely into mathematics (modulo arithmetic)
- Let us use the following transformation: a➛0, b➛1, c➛2, …, z➛25
- Then, the encryption of plaintext p with key k is E(k,p) = (p+k) mod 26
- Similarly, the decryption function is D(k,p) = (p-k) mod 26

Cryptanalysis of the shift cipher

Brute-force

- This is possible because of the small keyspace
- Try all 25 possible keys and choose the decryption that makes sense
- **Given ciphertext FUBSWRJUDSKB**
  - Key 1: etarvqitcrja
  - Key 2: dszquphsbqiz
  - **Key 3: cryptography**

Statistical attack

- Leverage the statistical properties of the plaintext
- In particular, the frequency of each character in the English language
- This can lead to an automated attack that does not require human supervision

English character frequencies (%)

a	b	c	d	e	f	g	h	i	j	k	l	m
8.2	1.5	2.8	4.2	12.7	2.2	2.0	6.1	7.0	0.1	0.8	4.0	2.4

n	o	p	q	r	s	t	u	v	w	x	y	z
6.7	7.5	1.9	0.1	6.0	6.3	9.0	2.8	1.0	2.4	0.2	2.0	0.1

Index of coincidence

- Given a text string, the index of coincidence (IOC) is the probability of two randomly selected characters being equal
- Let's associate the letters of the English alphabet with the numbers 0, 1, 2, 3, …, 25 (as mentioned before)
- Let p denote the vector of probabilities for a sufficiently large plaintext
- This is the vector shown on slide 6, i.e., $p_0 = 0.082$ (probability of character a)
- The IOC is given below and, for English text, it is approximately equal to 0.065 (for other languages it will be different)

$$I(p,p) = \sum_{i=0}^{25} p_i^2 \approx 0.065$$

A statistical attack on the shift cipher

- Given a ciphertext, let q denote the probability vector for the individual characters
- For example, q_0 denotes the probability of A, q_1 the probability of B, etc.
- Then, for every possible key value j (from 0 to 25), we compute the IOC between vectors p and a shifted version of vector q (by j positions)
- Specifically, we compute the following values:

$$I_j = I(p, q_{(i+j) \bmod 26}) = \sum_{i=0}^{25} p_i \cdot q_{(i+j) \bmod 26}$$

- If the key value is k, then I_k is approximately equal to 0.065
- **Attack:** Compute I_j for all j, and then output key value k for which I_k is closest to 0.065

Mono-alphabetic substitution cipher

- One problem with the shift cipher is that the shift value is the same for all plaintext characters
- This is why the keyspace is very small
- The idea behind mono-alphabetic substitution is to map each plaintext character to a different ciphertext character in an arbitrary manner
- In this case, the key is a table of one-to-one mapping, as in the example below

a	b	c	d	e	f	g	h	i	j	k	l	m
X	E	U	A	D	N	B	K	V	M	R	O	C

n	o	p	q	r	s	t	u	v	w	x	y	z
Q	F	S	Y	H	W	G	L	Z	I	J	P	T

Cryptanalysis of mono-alphabetic substitution

- The previous key will encrypt the message **attacknow** to **XGGXURQFI**
- This cipher has a very large keyspace: it consists of all permutations of the alphabet
- The number of permutations is $26! \approx 2^{88}$, so brute-force attacks are infeasible
- However, the mono-alphabetic substitution cipher is completely insecure
- We can use n-gram statistics to recover the substitution table:
  - Frequency of individual characters (1-gram)
  - Frequency of digrams (2-gram, two successive characters)
  - Frequency of trigrams (3-gram, three successive characters)
  - and so on
- The attack might still require some human supervision