1 – Introduction to Cryptography

Eklavya Sharma

Cryptography: The study of mathematical techniques for securing digital information, systems, and distributed computations against adversarial attacks.

Unlike modern cryptography, classical cryptography was based on ad-hoc techniques and lacked rigor.

Contents

1 Private-key encryption

Alice and Bob want to communicate over a public channel, but want to keep their communication private from an eavesdropper. Alice and Bob share a secret key k.

A private-key encryption scheme is the tuple $(\mathcal{M}, \mathcal{K}, \mathcal{C}, \mathsf{Gen}, e, d)$ where:

- $\mathcal{M}$ is a set called 'message space'.
- $\mathcal{K}$ is a set called 'key space'.
- $\mathcal{C}$ is a set called 'cipher space'.
- The key-generation algorithm Gen is a probabilistic algorithm that samples an item from $\mathcal{K}$.

- The encryption algorithm: For $k \in \mathcal{K}$, $e_k : \mathcal{M} \mapsto \mathcal{C}$.
- The decryption algorithm: For $k \in \mathcal{K}$, $d_k = e_k^{-1}$.

2 Kerckhoffs' Principle

Kerckhoffs says that the key must be secret, but the encryption scheme should be public.

Disadvantages of requiring the encryption scheme to be secret:

- Every pair of users will need a new algorithm.
- If the algorithm is leaked or lost, a new one will have to be invented.
- An algorithm which hasn't undergone public scrutiny is insecure.

3 Historical ciphers

All historical ciphers discussed here operate on strings of English characters. We denote the characters by $\Sigma = \mathbb{Z}_{26}$. $\mathcal{M} = \mathcal{C} = \Sigma^*$.

3.1 Caesar Cipher

- $\mathcal{K} = \{\}$
- $e_k(x)[i] = (x[i] + 3)\%26$
- $d_k(x)[i] = (x[i] - 3)\%26$

Trivial to break, since there is no key.

3.2 Shift Cipher

- $\mathcal{K} = \mathbb{Z}_{26}$; $|\mathcal{K}| = 26$.
- $e_k(x)[i] = (x[i] + k)\%26$
- $d_k(x)[i] = (x[i] - k)\%26$

Easy to break by brute force since key space is small. Frequency analysis will make it easier to guess which key to start with.

3.3 Mono-alphabetic Substitution Cipher

- $\mathcal{K} \in S_\Sigma$ ($\mathcal{K}$ is a permutation of Σ); $|\mathcal{K}| = 26! \approx 4.03 \times 10^{26}$.
- $e_k(x)[i] = k(x[i])$
- $e_d(x)[i] = k^{-1}(x[i])$

Can be broken using frequency analysis of the message space.

3.4 Vigenère Cipher

Also known as Poly-alphabetic shift cipher.

- $\mathcal{K} = \Sigma^*$.
- $e_k(x)[i] = (x[i] + k[i\%|k|])\%26$.
- $d_k(x)[i] = (x[i] - k[i\%|k|])\%26$.

If the key length is known, it can be broken using frequency analysis of every stream. For the cipher-text c and key-length l, the i^{th} stream is the sequence `c[i::l]` (python slice notation).

The key length can also be guessed using frequency analysis, like Kasisiki's method or using mean-square-frequency.

4 Definition of security

To mathematically prove that a cryptographic protocol is secure, we have to formally define what we mean by security.

There can be multiple definitions of security depending on the application and environment. Before developing a cryptographic solution to a problem, we must choose the definition of security that is most relevant to the application.

A security definition has 2 components: a security guarantee and a threat model.

Security guarantee for secure communication: Regardless of any information an attacker already has, a ciphertext should leak no additional information about the underlying plaintext.

4.1 Standard threat models for secure communication

- **Ciphertext-only attack**: Given $c = e_k(m)$ find out something about m.
- **Known-plaintext attack**: Given $T = \{(m, e_k(m)) : m \in S\}$ and $c = e_k(m)$ find out something about m.

- **Chosen-plaintext attack**: Given $c = e_k(m)$ and black-box access to e_k, find out something about m.

- **Chosen-ciphertext attack**: Given $c = e_k(m)$ and black-box access to e_k and d_k, find out something about m. The attacker is not allowed to feed c to d_k.