**EE 720: An Introduction to Number Theory and Cryptography (Spring 2019)**

Lecture 2 — January 7, 2019

*Instructor: Saravanan Vijayakumaran* *Scribe: Saravanan Vijayakumaran*

# 1 Lecture Plan

- Describe the difference between modern and classical cryptography
- Describe the syntax of private-key encryption
- Describe some classical ciphers – Caesar cipher, Substitution cipher, Vigenére cipher

# 2 Cryptography

- The dictionary definition of cryptography is "the art of writing or solving codes".
- Modern definition: The study of mathematical techniques for securing digital information, systems, and distributed computations against adversarial attacks.
- Modern approach to cryptography
  - Formal definitions
  - Precise assumptions
  - Proofs of security
- Private-key encryption setting
  - Also called symmetric-key setting
  - Communicating parties could be separated in space or time
- Syntax of encryption: Message space $\mathcal{M}$, Key generation procedure `Gen`, Encryption procedure `Enc`, Decryption procedure `Dec`
  - `Gen` is a probabilistic algorithm which generate key $k$
  - `Enc` takes $k$ and *plaintext* $m$ and gives *ciphertext* $c$ (probabilistic algorithm)
  - `Dec` takes $c$ and $k$ and gives $m$
- Kerckhoffs' principle: *The cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience.*
  - Security relies solely on secrecy of key
  - Easier to keep a short key secret than to keep an algorithm secret
  - Easier to change key than encryption scheme
  - Standardization is easier

# 3 Historical Ciphers

## 3.1 Caesar's cipher

- Always shift by 3. So effectively no key
- A variant called ROT-13 (shift by 13) is still used in social media.

## 3.2 Shift cipher

- Keyed variant of Caesar's cipher
- Key $k$ is a number between 0 and 25. `Enc` forward shifts by $k$ places. `Dec` backward shifts by $k$ places.
- Insecure cipher as there are only 26 possible keys, 25 if you ignore 0. Try decrypting using all keys and choose the decrypted output which looks like plaintext. Such an attack is called a brute-force or exhaustive-search attack.
- Sufficient key space principle: *Any secure encryption scheme should have a key space that is sufficiently large to make an exhaustive-search attack infeasible.*
- What is considered "infeasible"? Key space should have size at least $2^{70}$.
- Above principle gives a necessary condition for security, but not a sufficient one.

## 3.3 Substitution cipher

| Plaintext | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ciphertext | x | e | u | a | d | n | b | k | v | m | r | o | c | q | f | s | y | h | w | g | l | z | i | j | p | t |

Table 1: Example of a bijection on lowercase English alphabets

- Key $k$ is a bijection from {a,b,...,z} to {a,b,...,z}.
- Punctuation is ignored.
- Key space is $26! \approx 2^{88}$. Nevertheless scheme is easy to break using letter frequencies.
- Attack tabulates letter frequencies in the ciphertext and compares it to known letter frequencies.

| Letter | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency (%) | 8.2 | 1.5 | 2.8 | 4.3 | 12.7 | 2.2 | 2.0 | 6.1 | 7.0 | 0.2 | 0.8 | 4.0 | 2.4 | 6.7 | 1.5 | 1.9 | 0.1 | 6.0 | 6.3 | 9.1 | 2.8 | 1.0 | 2.4 | 0.2 | 2.0 | 0.1 |

Table 2: Average letter frequencies for English language text

### 3.4 Vigenére cipher

- Devised in the 16th century. Systematic attacks devised hundreds of years later.
- Assume a plaintext message consisting of a string of English alphabets without punctuation or spaces.
- The key is an English word of length $t$. It is repeated to create a alphabet string as long as the message.
- The message and the key are added modulo 26 to get the ciphertext.
- Key space has size $26^t$.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plaintext | `meetmetoday` | [12 | 04 | 04 | 19 | 12 | 04 | 19 | 14 | 03 | 00 | 24] |
| Key (repeated) | `peacepeacep` | [15 | 04 | 00 | 02 | 04 | 15 | 04 | 00 | 02 | 04 | 15] |
| Ciphertext | `bievqtxofen` | [01 | 08 | 04 | 21 | 16 | 19 | 23 | 14 | 05 | 04 | 13] |

Table 3: Example of Vigenére cipher operation

## 4 Principles of Modern Cryptography

### 4.1 Formal Definitions

- Formal definitions are needed to pin down exactly what "secure" means.
- Needed for the proper design, study, evaluation, and usage of cryptographic primitives.
- For example, consider secure encryption. What should a secure encryption scheme guarantee?
  - It should be impossible for an attacker to recover the key.
  - It should be impossible for an attacker to recover the entire plaintext from the ciphertext.
  - It should be impossible for an attacker to recover any character of the plaintext from the ciphertext.
  - *The right answer (informally)*: Regardless of any information an attacker already has, a ciphertext should leak no additional information about the underlying plaintext. Precise mathematical definition in later lectures

### 4.2 Threat Models

Threat models specify the abilities of the attacker but place no restrictions about the attacker's strategy, i.e. how the attacker uses those abilities.

There are several plausible threat models in the context of encryption. The following are the standard ones in increasing order of the attacker's ability.

- **Ciphertext-only attack**

  Adversary observes a ciphertext and attempts to determine information about the underlying plaintext

- **Known-plaintext attack**

  Adversary knows one or more plaintext/ciphertext pairs generated using the same key. Aims to deduce information about the underlying plaintext of some *other* ciphertext produced using the same key.

- **Chosen-plaintext attack**

  Similiar to above situtation, except that adversary can obtain plaintext/ciphertext pairs for plaintexts of its choice.

- **Chosen-ciphertext attack**

  Adversary is able to obtain the decryption of ciphertexts of its choice. Aims to deduce information about the underlying plaintext of some *other* ciphertext produced using the same key.

### 4.3 Precise Assumptions

Most modern cryptographic constructions cannot be proven secure unconditionally. Proofs of security rely on assumptions which need to made explicit and mathematically precise.

### 4.4 Proofs of Security

Proofs of security provide security guarantees relative to the definition being considered the specified assumptions being used.

## 5 Perfectly Secret Encryption

- Let us look at encryption schemes that are provably secure even against an adversary with unbounded computational power. Such schemes are called *perfectly secret*. The existence of such schemes is not obvious because we are allowing the adversary to launch brute-force attacks (for e.g., try all possible keys for any key length).
- This work was done by Shannon in the 1940s, so not exactly modern cryptography which is post 1970s. But Shannon was way ahead of his time.
- Recall the syntax of encryption: $m \in \mathcal{M}, k \in \mathcal{K}, k = \texttt{Gen}, c = \texttt{Enc}_k(m), m = \texttt{Dec}_k(c)$
- $c \leftarrow \texttt{Enc}_k(m)$ may be probabilistic but $\texttt{Dec}_k(c) = m$ with probability 1. This is called perfect correctness.
- Let $M$ be a random variable denoting the message (plaintext) being encrypted.
- Let $K$ be a random variable denoting the value of the key output by `Gen`. Almost always a uniform random variable on $\mathcal{K}$.

- $K$ and $M$ are assumed to be independent.
- Let $C$ be a random variable denoting the ciphertext.
- Fixing an encryption scheme and a distribution over $\mathcal{M}$ determines a distribution over $\mathcal{C}$ given by choosing a key $k \in \mathcal{K}$.
- Example: Consider the shift cipher with message set $\mathcal{M} = \{\texttt{kim}, \texttt{ann}, \texttt{boo}\}$ with probabilities 0.5, 0.2, 0.3 respectively. What is $\Pr[C = \texttt{dqq}]$? What is $\Pr[M = \texttt{ann} \mid C = \texttt{dqq}]$?

# 6 References and Additional Reading

- Preface and Sections 1.1, 1.2 from Katz/Lindell
- Read about the attacks on the Vigenére cipher in Section 1.3 from Katz/Lindell
- Sections 1.4, 2.1 from Katz/Lindell