Computer Security and Privacy January 24, 2019
Ashoka University CS 302
Debayan Gupta (with Barun Parruck and Rachit Rawat) Lecture 2

Symmetric Encryption Basics

–draft; to be updated at the end of the week–

A symmetric cryptosystem comprises:

- a set M of plaintext messages,
- a set C of ciphertexts,
- a set K of keys,
- an encryption function $E : K \times M \to C$
- a decryption function $D : K \times C \to M$

We often write $E_k(m) = E(k, m)$ and $D_k(c) = D(k, c)$.

You'll want to think about complexity and difficulty of implementation. Note that "k remains secret" is a naive claim of security; it is both too strong and too weak: Eve at always brute-force guess k, and you could have a cryptosystem where k remains private but m can be found easily.

Some attack scenarios (we'll discuss these later):

- Ciphertext only: Eve knows only c and tries to recover m.
- Known plaintext: Eve knows c and a set of plaintext-ciphertext pairs $(m_1, c_1), \ldots, (m_n, c_n)$ $c \notin \{c_1, \ldots, c_n\}$. Eve tries to recover m. (This might happen if the plaintext is revealed publicly later, such as in an auction, or be guessable, like an email header, etc.)
- Chosen plaintext
- Chosen ciphertext
- Both CPA and CCA
- Adaptive CPA, CCA

Think about when CPA and CCA would actually occur in real life. For now (though we might edit this later), we'll make do with this definition:

A cryptosystem is computationally secure relative to a notion of compromise if, for all probabilistic polynomial-time algorithms A, when given as input the "security parameter" n and all of the information available to Eve, the algorithm succeeds in compromising the cryptosystem with success probability that is negligible in n.

1 Some Definitions

You might be asked some questions about this in the exam. For the purposes of this lecture, assume that we're dealing with messages written in English.

1.1 Confusion and Diffusion

Many affect one; one affects many. Alternatively, one depends upon many; many depend upon one. Each item in the output should be sensitive to many things in the input; each item in the input should affect many things in the output.

1.2 Substitution Ciphers

Exactly what it sounds like. These cryptosystems replace each letter in the original message with a different letter. (Enigma reference: you might want to keep all the letters in play and not require that the letter be a different one – this was one of the things that helped break the Enigma machine.) In some cases, you might replace the letters with something from a different alphabet, but we won't discuss that here.

Any cipher that encrypts a message by applying the same substitution to each letter of the message is called a monoalphabetic cipher. Shift/Rotation cipher uses a such a substitution defined by a letter of the alphabet. Notice how a monoalphabetic cipher is much weaker (26 possibilities) than a full substitution cipher (26!).

1.3 Polyalphabetic Ciphers

Different letters can have different substitutions; essentially r different substitution ciphers used on consecutive letters. Permutation 1 is used on letter 1, permutation 2 on letter 2, ..., permutation r on letter r, permutation 1 on letter $r + 1$, etc. Any letter-level frequency attack usable on substitution ciphers can also be modified to attack these cryptosystems.

1.4 Polygraphic Ciphers

Encrypts several letters at a time. (The Hill cipher, discussed later in the class, is an example of this.)

1.5 Transposition Ciphers

These involve reorganizing the input in some way. E.g., you could switch the order of every pair of letters.

Rail fence / Scytale (write the text while moving up and down within a given height and read horizontally); route ciphers (write the text is some grid, then obey some rule or route when reading it).

Columnar transpostions would simply reorder the columns in a grid before reading them off row by row (the reordering could be based on a key; multiple such keys could be used to prevent anagramming).

Think about why grilles or cardboard cutout systems fall in this block. Also consider combination ciphers – e.g., combining substitution with a columnar transposition.

1.6 Fractionation

Each input letter is broken into multiple output or ciphertext letters/symbols. Transposing a fractionated input helps with diffusion.

1.7 Perfect Secrecy

Mentioned in the last lecture. Keep this in the back of your head.

2 Old Cryptography

2.1 Mesopotamia

Old clay tablets with pictographic representations are common enough, but some have deeper secrets!

2.2 Greece

Polybius squares used a grid to represent each letter by its coordinates in the grid (one could use some permutation of the alphabet in the grid though they didn't often do this) and then use these coordinates to substitute for the letter. Spartan Scytales (wrapping a message around a stick). Aeneas Tacticus wrote about some cryptographic techniques in his book on war.

2.3 Assyria/Babylon

Atbash Cipher – substitution with reversed alphabet (z is a etc.).

2.4 Egypt

Used to add extra meaning to some hieroglyphics; not particularly useful, as far was we can tell.

2.5 India

म्लेच्छित विकल्प "Mlecchita Vikalpa" in the Kama-Sutra. It is the 44th of the 64 arts, described in more detail in the (much later) Jayamangala.

Rebus-metonymy: Tagaraka and bowl-rim style symbolic writing (similar to some Chinese methods) in some very ancient tablets.

3 Caesar Cipher

Substitution cipher based on a simple rotation of the alphabet. HAL = IBM using the Augustus cipher (Caesar cipher with key 1).

4 Affine Cipher

Generalization of simple shift ciphers. Key is $\alpha, \beta \in \mathbb{Z}_{26}, gcd(\alpha, 26) = 1$. This effectively means that $\alpha \in \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$. We will be computing everything modulo 26, so there are only 26 unique values for β.

Encryption: $c = E_k(m) = \alpha m + \beta \mod 26$
Decryption: $D_k(m) = \alpha^{-1}(c - \beta) \mod 26$, where α^{-1} is the multiplicative inverse of α in the ring of integers $\mathbb{Z}_{26}$.

5 Vigenere Cipher

Uses a password where each letter defines a shift; this password is then repeated. This is a simple polyalphabetic cipher where the key is defined by a word or phrase.

You could improve this by using a nonrepeating message as the key, or perhaps parts of the message itself as an "autokey". All of these, however, can be broken with relative ease using some cleverness, frequency analysis, and modern computers.

6 Hill Cipher

This is a polygraphic cipher based on simple linear algebra. We use a non-singular square matrix and multiply each chunk of the input by this matrix (say, a 3×3 would be multiplied by chunks of size 3). Decryption would use the inverse of the matrix. A known-plaintext attack can easily break this particular cryptosystem.

7 Playfair Cipher

Invented by Wheatstone; this is a simple digraph cipher, where each pair of letters is replaced by some other pair, based on a key-grid. Each pair defines the corners of a rectangle in the grid and we use the letters at the other ends of the rows as substitutes, with some special rules in case both are in the same row or column (substitution is next letter, wrapping around if needed) or in the case of repeated letters (break it up by inserting X).

8 ADFGVX cipher

Used in WW1. Polybius square using ADFGX or ADFGVX (for all letters plus digits). This fractionated message is then put through a columnar transposition based on a keyword.

9 Vernam Cipher and One-Time Pads

Vernam created many variants of his system, but the one of most interest to us is the one-time pad. (You might also want to look into the

10 Block Ciphers

Next week