Fundamentals of classical cryptosystems & one-time pads

COSC312 + COSC412—Cryptography and Security

David Eyers (with thanks to Michael Albert!)

Basic problem

- **Alice** wishes to send **Bob** a confidential message whose contents may be of interest to a third party, **Eve**.
  - What resources can Eve use to discover message contents?

- **Objective**: It should be at least as difficult for Eve to reconstruct the message having intercepted it, as it would be to suborn the process in some other way.
  - *I.e., message security* is no worse than *general security*.

Messages and keys

- **Message space**, $\mathcal{M}$: set of all possible messages
  - Messages are just sequences of bits (or bytes / words / strings).
- A **key** is a piece of genuinely private information held by Alice and Bob (but not Eve!).
- The **key space**, $\mathcal{K}$, is the set of all possible keys.

Symmmetric cryptosystems

- A **symmetric cryptosystem** is a pair of functions:

$$E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$$
$$D : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$$

such that for all $m \in \mathcal{M}$ and all $k \in \mathcal{K}$,

$$D(k, E(k, m)) = m.$$

i.e., if from message m, Alice produces ciphertext $c = E(k, m)$, Bob can recover $m = D(k, c)$.

Let's discuss four types of attack

- **Ciphertext-only attack**: Eve has access only to the encrypted message c.
  - (. . . or possibly some sequence of encrypted messages.)
- **Known-plaintext attack**: Eve has access to some pairs (m, c) of previous messages and ciphertexts.
- **Chosen-plaintext attack**: Eve can choose plaintexts and obtain the corresponding ciphertexts.
- **Brute-force attack**: Eve can try all possible keys.

Caesar cipher (circa 58BCE)

- Take $\mathcal{M}$ to be the space of strings over $\mathcal{A}$, the set of upper case letters, A through Z.
  - Think of the letters as numbers: $A \leftrightarrow 0$ through $Z \leftrightarrow 25$.
- Take $\mathcal{K}$ to be the set of upper case letters, and let k be a particular key.
- E just "adds k" to each letter of the message (wrapping around, *i.e.*, taking a remainder modulo 26).
- D just "subtracts k" (wrapping around as needed).

Substitution cipher

- Take $\mathcal{M}$ to be the space of strings of upper case letters, *A* through *Z*.
- Take $\mathcal{K}$ to be the set of permutations of $\mathcal{A}$, and let κ be a particular key.
- *E* just applies κ to each letter of the message.
- *D* just applies the inverse of κ

Vigenère cipher (sixteenth century)

- Take $\mathcal{M}$ to be the space of strings of upper case letters, A through Z.
- Take $\mathcal{K}$ to be the set of strings from $\mathcal{A}$ of some fixed length, n, and let $\mathbf{k} = k_0 k_1 k_2 \ldots k_{n-1}$ be a particular key.
- E applies the Caesar ciphers corresponding to the characters of $\mathbf{k}$ sequentially to characters of m, wrapping back to the beginning of $\mathbf{k}$ when necessary.

Breaking Vigenère ciphers 1/2

- To break the Vigenère cipher it's pretty much sufficient to be able to **work out the key length**.

- **Friedman test** steps:
  - break up the text according to an assumed key length;
  - if correct, each block will represent a sample of letters according to the standard frequency distribution (rotated);
  - if incorrect, each block will represent a mixture of two or more such samples (with different rotations) so will be 'smoother';
  - try to quantify smoothness over tested key lengths.

Breaking Vigenère ciphers 2/2

- **Kasiski examination** can be a powerful technique:
  - Look for **repeated bigrams** or trigrams in ciphertext;
  - **gaps between** them are likely to be multiples of the **key length**.

- Questions to consider:
  - Why does the Kasiski examination work?
  - What difficulties might arise in practice?

Key insights about attacks so far

- Ciphertext only attacks on classical cryptosystems are based on discovering **patterns in the ciphertext** that correspond to the structure of the plaintext.
- Fundamental goal: discover **information about the key**
  - Ideally discover enough so **brute force** can finish off the job.
  - As computing resources increase these attacks grow stronger.
- Any cryptosystem which creates such patterns must be **deemed to be (potentially) insecure**.

Avoiding patterns in ciphertext

- Observation: **random text** contains no patterns… so…
- Can we create cryptosystems in which the ciphertext is—or appears to be—random and yet still contains the information we want to transmit?
  - …and if so, how do we achieve this?

Leaving Eve in the dark (perfect secrecy)

- Claude Shannon is a key figure in information theory.
- He observed about cryptosystems that if,
  - for any two messages m_0 and m_1 and any ciphertext c,
  - the number of keys, k, such that $E(k, m_0) = c$ is the same as the number of keys k' such that $E(k', m_1) = c$, and
  - keys are chosen uniformly at random.
- Then the ciphertext alone contains **no information** about the message—the cipher has **perfect secrecy**.
  - . . . also, no ciphertext-only attacks!

Perfect secrecy in context; also ‘exclusive or’

- Our Vigenère cipher use doesn’t give perfect secrecy
  - How can we demonstrate this?
- Can we achieve perfect secrecy?
- Aside: recall ‘exclusive or’ (also ‘XOR’ or $\oplus$) operation.
  - For two binary digits (bits) A and B, $\oplus$ gives:

A	B	$A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

Introducing the one-time pad

- Two same-length bit-strings a and b, $\oplus$ is bit-by-bit XOR.

$$\mathbf{2} = \{0, 1\} \qquad \text{the set of } \textit{bits}$$
$$\mathcal{M} = \mathcal{K} = \mathbf{2}^n \qquad \text{the set of } n\text{-bit strings}$$
$$E(k, m) = k \oplus m \qquad \text{encoding}$$
$$D(k, c) = k \oplus c \qquad \text{decoding}$$

- For any message m and any ciphertext c there's exactly one k such that $E(k, m) = c$. In fact, $k = c \oplus m$.
  - So it **achieves perfect secrecy**! But is it practical?

Problems with one-time pads

- Approach does not preserve **message integrity**
  - *i.e.*, if Eve can guess structure, she can modify part of it.
  - However, we can protect against this fairly easily.
- More significantly, **key is same size as the message** and this is necessary for perfect secrecy (why?)
  - If Alice and Bob can agree on that secret key, then why don't they just spend that time communicating the message?
- We'll consider consequences of the above next lecture
  - …and methods to deal with the issues raised