Introduction to Cryptology

Lecture 3

Announcements

- Homework 1 due on Thursday 2/5
  - Hand in code for Problem 1 and decrypted ciphertext along with other solutions
- Readings up on course webpage on Computational Complexity
  - We will start computational approach to cryptography next week.

Agenda

- Last time:
  - Cryptanalysis of the Vigenere Cipher (1.3)
  - Terminology and Definitions
- This time:
  - Terminology and Definitions
  - Formal definition of a symmetric encryption scheme (2.1)
  - Shannon's definition of perfect secrecy (2.1)
  - Equivalent definitions (2.1)
  - Construction of a perfectly secret scheme (2.2)

Terminology

- Discrete Random Variable: A discrete random variable is a variable that can take on a value from a finite set of possible different values each with an associated probability.
- Example: Bag with red, blue, yellow marbles. Random variable X describes the outcome of a random draw from the bag. The value of X can be either red, blue or yellow, each with some probability.

More Terminology

- **A discrete probability distribution** assigns a probability to each possible outcomes of a discrete random variable.
  - Ex: Bag with red, blue, yellow marbles.
- An **experiment** or **trial** (see below) is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space.
  - Ex: Drawing a marble at random from the bag.
- An **event** is a set of outcomes of an experiment (a subset of the sample space) to which a probability is assigned
  - Ex: A red marble is drawn.
  - Ex: A red or yellow marble is drawn.

Conditional Probability

- A **conditional probability** measures the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.
- Probability of event X, conditioned on event Y: $\Pr[X \mid Y]$
- Example: Probability the second marble drawn will be red, conditioned on the first marble being yellow.

Basic Facts from Probability

- If two events are independent if and only if $\Pr[X \mid Y] = \Pr[X]$.
- AND of two events: $\Pr[X \wedge Y] = \Pr[X] \cdot \Pr[Y \mid X]$
- AND of two independent events: $\Pr[X \wedge Y] = \Pr[X] \cdot \Pr[Y]$
- OR of two events: $\Pr[X \vee Y] \leq \Pr[X] + \Pr[Y]$
  - This is called a "union bound."

Formally Defining a Symmetric Key Encryption Scheme

Syntax

- An encryption scheme is defined by three algorithms
  - Gen, Enc, Dec
- Specification of message space $\boldsymbol{M}$ with $|\boldsymbol{M}| > 1$.
- Key-generation algorithm Gen:
  - Probabilistic algorithm
  - Outputs a key k according to some distribution.
  - Keyspace $\boldsymbol{K}$ is the set of all possible keys
- Encryption algorithm Enc:
  - Takes as input key $k \in \boldsymbol{K}$, message $m \in \boldsymbol{M}$
  - Encryption algorithm may be probabilistic
  - Outputs ciphertext $c \leftarrow Enc_k(m)$
  - Ciphertext space $\boldsymbol{C}$ is the set of all possible ciphertexts
- Decryption algorithm Dec:
  - Takes as input key $k \in \boldsymbol{K}$, ciphertext $c \in \boldsymbol{C}$
  - Decryption is deterministic
  - Outputs message $m \coloneqq Dec\_k(c)$

Distributions over K, M, C

- Distribution over $\boldsymbol{K}$ is defined by running Gen and taking the output.
  - For $k \in \boldsymbol{K}, \Pr[K = k]$ denotes the prob that the key output by Gen is equal to k.
- For $m \in \boldsymbol{M}, \Pr[M = m]$ denotes the prob. That the message is equal to m.
  - Models a priori knowledge of adversary about the message.
  - E.g. Message is English text.
- Distributions over $\boldsymbol{K}$ and $\boldsymbol{M}$ are independent.
- For $c \in \boldsymbol{C}, \Pr[C = c]$ denotes the probability that the ciphertext is c.
  - Given Enc, distribution over $\boldsymbol{C}$ is fully determined by the distributions over $\boldsymbol{K}$ and $\boldsymbol{M}$.

Definition of Perfect Secrecy

- An encryption scheme (Gen, Enc, Dec) over a message space $\boldsymbol{M}$ is perfectly secret if for every probability distribution over $\boldsymbol{M}$, every message $m \in \boldsymbol{M}$, and every ciphertext $c \in \boldsymbol{C}$ for which $\Pr[C = c] > 0$:

$$\Pr[M = m \mid C = c] = \Pr[M = m].$$

An Equivalent Formulation

- Lemma: An encryption scheme (Gen, Enc, Dec) over a message space $\boldsymbol{M}$ is perfectly secret if and only if for every probability distribution over $\boldsymbol{M}$, every message $m \in \boldsymbol{M}$, and every ciphertext $c \in \boldsymbol{C}$:
$$\Pr[C = c \mid M = m] = \Pr[C = c].$$

Basic Logic

- Usually want to prove statements like $P \to Q$ ("if P then Q")
- To prove a statement $P \to Q$ we may:
  - Assume P is true and show that Q is true.
  - Prove the contrapositive: Assume that Q is false and show that P is false.

Basic Logic

- Consider a statement $P \leftrightarrow Q$ (P if and only if Q)
  - Ex: Two events X, Y are independent if and only if $\Pr[X \wedge Y] = \Pr[X] \cdot \Pr[Y]$.
- To prove a statement $P \leftrightarrow Q$ it is sufficient to prove:
  - $P \rightarrow Q$
  - $Q \rightarrow P$

Proof (Preliminaries)

- Recall Bayes' Theorem:
  - $\Pr[A \mid B] = \frac{\Pr[B|A]\cdot\Pr[A]}{\Pr[B]}$

- We will use it in the following way:
  - $\Pr[M = m \mid C = c] = \frac{\Pr[C=c \mid M=m]\cdot\Pr[M=m]}{\Pr[C=c]}$

Proof

Proof: $\rightarrow$

- To prove: If an encryption scheme is perfectly secret then

"for every probability distribution over $\boldsymbol{M}$, every message $m \in \boldsymbol{M}$, and every ciphertext $c \in \boldsymbol{C}$:

$$\Pr[C = c \,| M = m] = \Pr[C = c]\,."$$

Proof (cont'd)

- Fix some probability distribution over $\boldsymbol{M}$, some message $m \in \boldsymbol{M}$, and some ciphertext $c \in \boldsymbol{C}$.
- By perfect secrecy we have that

$$\Pr[M = m \mid C = c] = \Pr[M = m].$$

- By Bayes' Theorem we have that:

$$\Pr[M = m \mid C = c] = \frac{\Pr[C = c \mid M = m] \cdot \Pr[M = m]}{\Pr[C = c]} = \Pr[M = m].$$

- Rearranging terms we have:

$$\Pr[C = c \mid M = m] = \Pr[C = c].$$