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 \( \Pr[X | Y] = \Pr[X] \).

• AND of two events: \( \Pr[X \land Y] = \Pr[X] \cdot \Pr[Y | X] \)

• AND of two independent events: \( \Pr[X \land Y] = \Pr[X] \cdot \Pr[Y] \)

• OR of two events: \( \Pr[X \lor Y] \leq \Pr[X] + \Pr[Y] \)
  – 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 $\mathcal{M}$ with $|\mathcal{M}| > 1$.
- Key-generation algorithm $Gen$:
  - Probabilistic algorithm
  - Outputs a key $k$ according to some distribution.
  - Keyspace $\mathcal{K}$ is the set of all possible keys
- Encryption algorithm $Enc$:
  - Takes as input key $k \in \mathcal{K}$, message $m \in \mathcal{M}$
  - Encryption algorithm may be probabilistic
  - Outputs ciphertext $c \leftarrow Enc_k(m)$
  - Ciphertext space $\mathcal{C}$ is the set of all possible ciphertexts
- Decryption algorithm $Dec$:
  - Takes as input key $k \in \mathcal{K}$, ciphertext $c \in \mathcal{C}$
  - Decryption is deterministic
  - Outputs message $m := Dec_k(c)$
Distributions over $K, M, C$

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

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

• Usually want to prove statements like \( P \rightarrow Q \) (“if \( P \) then \( Q \”\)

• To prove a statement \( P \rightarrow Q \) 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 $P \iff Q$ ($P$ if and only if $Q$)
  – Ex: Two events $X, Y$ are independent if and only if $\Pr[X \land Y] = \Pr[X] \cdot \Pr[Y]$.
• To prove a statement $P \iff Q$ it is sufficient to prove:
  – $P \implies Q$
  – $Q \implies P$
Proof (Preliminaries)

• Recall Bayes’ Theorem:

\[
- \Pr[A \mid B] = \frac{\Pr[B \mid A] \cdot \Pr[A]}{\Pr[B]}
\]

• 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: →

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

“for every probability distribution over $M$, every message $m \in M$, and every ciphertext $c \in C$:

$\Pr[C = c \mid M = m] = \Pr[C = c]$. "
Proof (cont’d)

• Fix some probability distribution over $M$, some message $m \in M$, and some ciphertext $c \in C$.

• 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:
  \[ \Pr[C = c \mid M = m] = \Pr[C = c]. \]