ECE596C: Handout #5

Elements of Shannon’s Approach to Cryptography

Electrical and Computer Engineering, University of Arizona,
Loukas Lazos

Abstract. In this lecture we present basic concepts of Information theory and how it can be applied to evaluate the performance of a cryptosystem. Readings from Chapter 2 of D. Stinson.

1 Entropy

Let $X$ be a random variable: $P(X = x) = p(x)$. Note that $\sum_{x \in X} p(x) = 1$. The binary Entropy of random variable $X$ is defined as:

$$H_2(X) = - \sum_{x \in X} p(x) \log_2 p(x) \text{ bits.} \quad (1)$$

As an example consider a coin toss. Let $P(H) = p$, and $P(T) = 1 - p$, such that $P(H) + P(T) = 1$. The entropy of the coin toss:

$$H_2(p) = -p \log_2 p - (1 - p) \log_2 (1 - p). \quad (2)$$

When $p = 0$: $H_2(p) = -0 \log_2 0 - 1 \log_2 1 = 0$.
When $p = 1$: $H_2(p) = -1 \log_2 1 - 0 \log_2 0 = 0$.

Without any proof we can infer that the entropy is concave in $p$ as shown in Figure 2, and has a maximum at $p = \frac{1}{2}$. The maximum entropy is $H_2(\frac{1}{2}) = -\frac{1}{2} \log_2 1 - \frac{1}{2} \log_2 \frac{1}{2} = 1\text{ bit}$. Hence the maximum entropy occurs when the coin is fair (when $p = \frac{1}{2}$).

As a philosophical insight, we can see from the experiment above, entropy has the value based on probability and is maximum when the coin is fair. But when the coin is fair we have the highest uncertainty about outcome of the experiment. Hence we could state that entropy is the amount of uncertainty in a random variable or experiment outcome.

2 Properties of Entropy

Definition 1. A real-valued function is said to be concave on an interval $I$ if:

$$f \left( \frac{x + y}{2} \right) \geq \frac{f(x) + f(y)}{2}, \quad (3)$$

$\forall x, y \in I$. The function $f$ is said to be strictly concave if a strict inequality holds for any $x \neq y$.

One of the properties of concave functions is that a local optimality guarantees global optimality. In the case where the function $f$ is twice differentiable, then the second derivative is negative. A very easy graphical test to verify the concavity of any function is draw any cord. A concave function should lie above the cord.
Theorem 1. Jensen’s Inequality: Suppose that $f$ is a continuous strictly concave function on interval $I$ and that

$$
\sum_{i=1}^{n} a_i = 1,
$$

with $a_i > 0, 1 \leq i \leq n$. Then,

$$
\sum_{i=1}^{n} a_i f(x_i) \leq f\left(\sum_{i=1}^{n} a_i x_i\right),
$$

where $x_i \in I, 1 \leq i \leq n$. Further equality holds if and only if $x_1 = x_2 = \cdots = c_n$.

We use Jensen’s inequality to prove that the entropy of a random variable $X$ attains its maximum value when $X$ is uniformly distributed.

Theorem 2. Let $X$ be a random variable which takes $n$ values with probabilities $p_1, p_2, \ldots, p_n$ with $p_i > 0, \forall i$. $H(X) \leq \log_2(n)$, with equality holding only when $p_i = \frac{1}{n}, \forall i$.

Proof.

$$
H(X) = -\sum_{i=1}^{n} p_i \log p_i
= \sum_{i=1}^{n} p_i \log \frac{1}{p_i}
\leq \log_2 \sum_{i=1}^{n} p_i \frac{1}{p_i}
= \log_2 n.
$$
with equality holding if and only if \( p_i = \frac{1}{n}, 1 \leq i \leq n \).

### 2.1 Joint Entropy and Conditional Entropy

We can extend the definition of entropy to multiple random variables, and consider the joint entropy. The definition for two random variables is as follows.

**Definition 2.** The joint entropy \( H(X, Y) \) of two discrete random variables \((X, Y)\) with a joint probability distribution \( p(x, y) \) is defined as

\[
H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y). \tag{7}
\]

Similarly, we can extend the definition for any number of random variables.

**Theorem 3.** \( H(X, Y) \leq H(X) + H(Y) \), with equality holding only if \( X, Y \) are independent random variables.

**Definition 3.** Conditional Entropy Let \( X, Y \) be two random variables. We can then define the conditional probability distribution of \( Y \) given \( x \) and denote the entropy associated with the variable \( Y|X = x \) by

\[
H(X|Y) = \sum_{y \in Y} p(y)H(X|Y = y)
= -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y)
= -\sum_{y \in Y} \sum_{x \in X} p(x, y) \log p(x|y).
\]

The conditional entropy is a measure of the average amount of information about \( X \), that is not revealed by the knowledge of \( Y \).

**Theorem 4.** \( H(X, Y) = H(Y) + H(X|Y) \).

**Corollary 1.** \( H(X|Y) \leq H(X) \), with equality holding if and only if \( X, Y \) are independent random variables.

### 2.2 Mutual Information

The “opposite” measure of conditional entropy can be consider to be the mutual information between two random variables \( X, Y \), which is the measure of the information revealed by the knowledge of a random variable \( X \), regarding the random variable \( Y \).

**Definition 4.** The mutual information for two random variables \( X, Y \) is defined as the relative entropy between the joint distribution \( p(x, y) \) and the product of the marginal distributions \( p(x), p(y) \).

\[
I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{8}
\]
2.3 Relationships between mutual Information and Entropy

\[
I(X; Y) = H(X) - H(X|Y), \quad (9)
\]
\[
I(X; Y) = H(Y) - H(Y|X), \quad (10)
\]
\[
I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (11)
\]
\[
I(X; Y) = I(Y; X), \quad (12)
\]
\[
I(X; X) = H(X). \quad (13)
\]

3 Application of Entropy to Cryptosystems

**Definition 5. Key Equivocation:** The amount of uncertainty of the key, when the ciphertext is known.

**Theorem 5.** Let \((P, K, C, E, D)\) be a cryptosystem. Then,

\[
H(K|C) = H(K) + H(P) - H(C). \quad (14)
\]

**Proof.** Based on the definition of conditional entropy:

\[
H(K, P, C) = H(C|K, P) + H(K, P) = H(K, P) = H(K) + H(P). \quad (15)
\]

![Relationships between conditional entropy and mutual information](image_url)
Similarly we can write

\[ H(K, P, C) = H(P|K, C) + H(K, C) = H(K, C). \]  

(16)

We can now compute

\[ H(K|C) = H(K, C) - H(C) \]
\[ = H(K, P, C) - H(C) \]
\[ = H(K) + H(P) - H(C). \]

To maximize the uncertainty of an adversary in determining key \( K \) we must maximize the conditional entropy \( H(K|C) \). But \( H(K|C) \leq H(K) \) with equality holding only when \( K, C \) are independent. For this to happen \( H(P) = H(C) \).

**Example:** Let \( P = \{a, b\} \), with \( \text{Pr}[a] = 0.25 \) and \( \text{Pr}[b] = 0.75 \). Let also \( K = \{K_1, K_2, K_3\} \) having probability distribution of 0.5, 0.25, 0.25, respectively. Let the ciphertext distribution be \( C = \{1, 2, 3, 4\} \) with the encryption function be given by the following matrix

\[
\begin{array}{ccc}
  a & b \\
  K_1 & 1 & 2 \\
  K_2 & 2 & 3 \\
  K_3 & 3 & 4 \\
\end{array}
\]

**Table 1.** The encryption matrix.

Given the probability distributions for \( P, K \) we have calculated \( H(P) \approx 0.85 \), \( H(K) \approx 1.5 \) and \( H(C) \approx 1.85 \). Based on the above theorem we can compute \( H(K|C) \approx 0.46 \). We can also compute \( H(K|C) \) analytically by using the probabilities \( \text{Pr}[K = k|Y = y] \).

\[
\begin{array}{c}
  \text{Pr}[K_1] = 1 \\
  \text{Pr}[K_2] = \frac{6}{7} \\
  \text{Pr}[K_3] = 0 \\
  \text{Pr}[K_1] = 0 \\
  \text{Pr}[K_2] = \frac{1}{7} \\
  \text{Pr}[K_3] = 0 \\
\end{array}
\]

We can now compute

\[ H(K|C) = \frac{1}{8} \times 0 + \frac{7}{16} \times 0.59 + \frac{1}{4} \times 0.81 + \frac{3}{16} \times 0 = 0.46. \]  

(17)

**Definition 6. Spurious Keys** The set of keys that decrypt the ciphertext giving plausible but incorrect plaintext.

**Example of a spurious key** Let the ciphertext be \( W.N.AJW \) and let the shift cipher be the cryptosystem used. Assuming that the cipher is the result of the encryption of a word of the English
language, decrypting the ciphertext with all possible key values we obtain the following two “meaningful” ciphertxts: arena, river, for encryption keys \( F = 5 \) and \( W = 22 \), respectively. Given that only one is the correct key, the other key is assumed to be spurious. If, for example \( F \) was the right key, then \( W \) is assumed to be a spurious key.

**Definition 7.** Suppose \( L \) is a natural language. The entropy of \( L \) is defined as

\[
H_L = \lim_{n \to \infty} \frac{H(P^n)}{n}
\]

and the redundancy of \( L \) is defined as

\[
R_L = 1 - \frac{H_L}{\log_2 |P|}.
\]

The entropy of a natural language measures the entropy of the language per letter. If the language consists of random letters then the entropy is \( \log_2 |P| \). The maximum entropy for the English alphabet is \( \log_2 26 \approx 4.70 \).

Entropy of monograms : \( H(P) \approx 4.19 \)
Entropy of digrams:: \( H(P^2) \approx 3.90 \)

Empirical studies have shown that the entropy of the English language is \( 1.0 \leq H_L \leq 1.5 \). Using 1.25 as an estimate we get \( R_L = 0.75 \), or that the English text is approximately 75% redundant! (i.e. One can compress the English text by 75%).

### 3.1 A bound on the expected number of spurious keys

Let \( C^n \) denote the random variable representing an n-gram of ciphertxt. For a \( y \in C^n \),

\[
K(y) = \{ K \in K : \exists x \in P^n \text{ such that } Pr[x] > 0 \text{ and } e_K(x) = y \}.
\]

That is, \( K(y) \) is the set of all keys for which the ciphertext \( y \) yields a meaningful string of plaintext of length \( n \). For each \( y \) the number of spurious keys is equal to \( |K(y)| - 1 \). We can compute the average number of spurious keys over all possible ciphertxts \( y \) of length \( n \) as

\[
\bar{s}_n = \sum_{y \in C^n} Pr[y] (|K(y)| - 1)
= \sum_{y \in C^n} Pr[y] |K(y)| - \sum_{y \in C^n} Pr[y]
= \sum_{y \in C^n} Pr[y] |K(y)| - 1.
\]

From a prior theorem we can compute

\[
H(K|C^n) = H(K) + H(P^n) - H(C^n).
\]

We can also estimate \( H(P^n) \approx nH_L = n(1 - R_L) \log_2 |P| \), given that \( n \) is reasonably large. We also have \( H(C^n) \leq n \log_2 |C| \). For cryptosystems where \( |C| = |P| \) it follows that
We can now relate $H(K|C^n)$ with $\bar{s}_n$.

\[
H(K|C^n) = \sum_{y \in C^n} \Pr[y] H(K|y) 
\leq \sum_{y \in C^n} \Pr[y] \log_2 |K(y)| 
\leq \log_2 \sum_{y \in C^n} \Pr[y]|K(y)| 
= \log_2(\bar{s}_n + 1),
\]  
(24)

where we have applied Jensen’s inequality. Combining the two inequality yields,

\[
\log_2(\bar{s}_n + 1) \geq H(K) - nR_L \log_2 |P|. \]  
(25)

If the keys are chosen equiprobably,

**Theorem 6.** For a cryptosystem with $|P| = |C|$ and keys chosen randomly (uniformly) from $K$, the expected number of spurious keys is lower bounded by:

\[
\bar{s}_n \geq \frac{|K|}{|P|^n R_L} - 1.
\]  
(26)

**Definition 8.** The unicity distance of a cryptosystem is defined to be the value $n_0$ for which the expected number of spurious keys becomes zero, i.e., the average amount of ciphertext required to uniquely define the key, given sufficient computation time.

We can estimate the unicity distance from the above theorem by setting $\bar{s}_n = 0$.

\[
n_0 \approx \frac{\log_2 |K|}{R_L \log_2 |P|}.
\]  
(27)

**Example:** For the substitution cipher we have $|P| = 26$ and $|K| = 26!$ If we consider $R_L = 0.75$ the unicity distance can be computed to be

\[
n_0 \approx \frac{88.4}{0.75 \times 4.7} \approx 25.
\]  
(28)