Shannon's Theory of Cryptography

CS702: Seminar

by R. Vijay Shankar (CS93136)

Instructor: Prof. C. Pandu Rangan

March 19, 1997

1 Introduction to Cryptosystems

The fundamental goal of cryptography is to allow *secure communication over an insecure channel*. Traditionally, this is illustrated as follows: Two people Alice and Bob want to communicate over a channel over which an opponent Oscar can eavesdrop. The message that Alice sends will be in *plaintext* i.e. English, French, a graph, or any form which can be understood by all. To prevent Oscar from reading this message, she encrypts this message into a *ciphertext* which will be decrypted by Bob at the other end. Oscar, who has access only to the ciphertext, cannot understand the message. This process is depicted in Figure 1.

Formally, we define a *cryptosystem* to be a 5-tuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ where $\mathcal{P}$ is a finite set of plaintexts, $\mathcal{C}$ is a finite set of ciphertexts and $\mathcal{K}$ is a set of keys. For each key $k \in \mathcal{K}$, there exists an *encryption rule* $e_k : \mathcal{P} \rightarrow \mathcal{C}$ and a corresponding *decryption rule* $d_k : \mathcal{C} \rightarrow \mathcal{P}$ where $e_k \in \mathcal{E}$ and $d_k \in \mathcal{D}$.

Each plaintext say x is encrypted by Alice with a key say k using e_k to produce a ciphertext y and sent to Bob, who decrypts y using the decryption function d_k. Therefore we must have $d_k(e_k(x)) = x \ \ \forall x \in \mathcal{P}$. Also, e_k must be an injective function for unambiguous decryption.

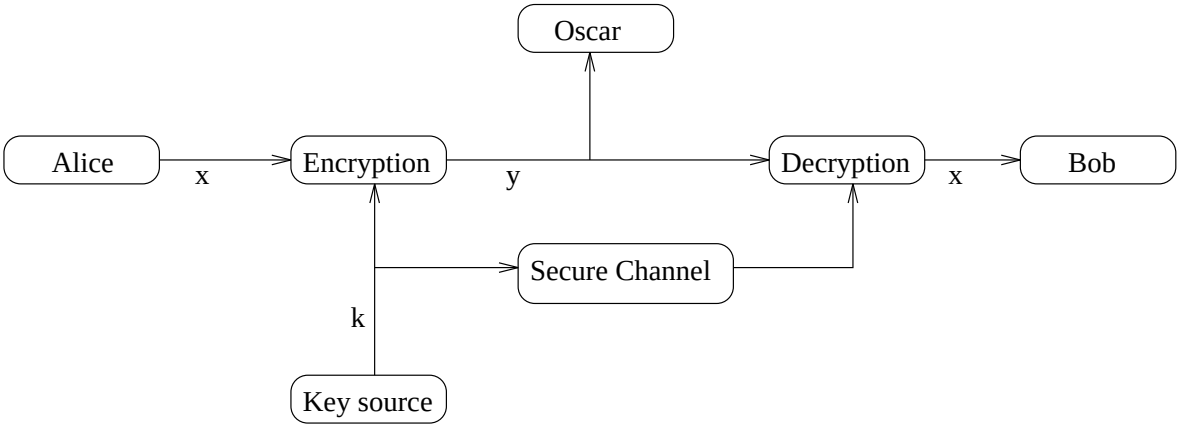

Figure 1: The Encryption-Decryption Process

Some Simple Cryptosystems

We now describe some simple cryptosystems and their encryption/decryption mechanisms.

The Shift Cipher

This is a very simple cipher where letters of the alphabet are encypted by shifting them by a constant number of positions. Formally we have,
$\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{Z}_{26}$, corresponding to letters of the Enlish alphabet.
For all k, $0 \leq k \leq 26$, define $e_k(x) = x+k \bmod 26$ and, $d_k(y) = y-k \bmod 26$ where $x, y \in \mathcal{Z}_{26}$. For example, if the plaintext is "wewillmeetatmidnight", then with the key 11, the ciphertext will be "hphtwwxppelextoytrse". This cipher is supposed to have been used by Julius Caesar for his military secrets.

The Substitution Cipher

Here, the ciphertext is formed from the plaintext by simply mapping each letter of the alphabet into another letter. Therefore, $\mathcal{P} = \mathcal{C} = \mathcal{Z}_{26}$, corresponding to letters of the Enlish alphabet and $\mathcal{K}$ is the set of all permutations over the

English alphabet. For each permutation $\pi \in \mathcal{K}$, we define $e_\pi(x) = \pi(x)$ and $d_\pi(y) = \pi^{-1}(y)$, where π^{-1} is the inverse permutation and $x, y \in \mathcal{Z}_{26}$.

2 Shannon's Theory of Cryptography

In 1949, Claude Shannon published an article on the "Communication Theory of Secrecy Systems" in the *Bell Systems Technical Journal*. This paper had a catalytic effect on the study of cryptography and it laid a formal, rigorous basis for the developmenet of modern cryptography. In the rest of this report, we shall discuss the main ideas in his work.

Types Of Security

Computational Security: Any algorithm for breaking the cryptosystem needs atleast a certain specified number of operations.

Unconditional Security: There is no way to break the cryptosystem, even with no bound on the amount of computation.

We develop the theory of *unconditional security* against a *ciphertext-only attack*. In other words, the opponent Oscar has access to the cipher-text and tries to find the key used. We now state an important assumption: *The opponent, Oscar, knows the cryptosystem being used. In other words, although he does not know the exact key, he knows the method used for encryption and decryption.*

Initially, we shall assume that *each key is used for only one encryption.* We give below the notation that will be used in the rest of this report.
Notation: $p_P(x)$ denotes the probability that a $x \in \mathcal{P}$ occurs and $p_K(k)$ gives the probability that a $k \in \mathcal{K}$ is chosen. $p_C(y)$ gives the probability that a $y \in \mathcal{C}$ results after encryption. We assume that p_P and p_K are independent. For each $k \in \mathcal{K}$, let $C(k) = \{e_k(x) : x \in \mathcal{P}\}$. $C(k)$ is the set of possible ciphertexts for key k.

Given the *a priori* probabilities $p_P(x)$ and $p_K(x)$, we can easily evaluate $p_C(y)$ as follows.

$$p_C(y) = \sum_{k:y\in C(k)} p_K(k)p_P(d_k(y)) \text{ and, } p_C(y|x) = \sum_{k:x=d_k(y)} p_K(k)$$

Therefore the *a posteriori* probability $p_P(x|y)$ is

$$p_P(x|y) = \frac{p_P(x)\sum_{k:x=d_k(y)} p_K(k)}{\sum_{k:y\in C(k)} p_K(k)p_P(d_k(y))}$$

3 Perfect Secrecy

Definition 1 *A Cryptosystem $(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{D},\mathcal{E})$ is defined to have* Perfect Secrecy *if $p_P(x|y) = p_P(x)$ $\forall x \in \mathcal{P}, y \in \mathcal{C}$. Intuitively, Perfect Secrecy implies that the ciphertext y gives no clue as to the plaintext.*

Theorem 1 *If each key is used with equal probability $\frac{1}{26}$ then the Shift Cipher over the English alphabet gives perfect secrecy for any probability distribution p_P.*

Proof: For the shift cipher, $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{Z}_{26}$. The encryption function is given by $e_k(x) = x + k \mod 26$. Thus $p_C(y) = \sum_{k\in\mathcal{Z}_{26}} p_K(k)p_P(d_k(y))$
$= \sum_{k\in\mathcal{Z}_{26}} p_P(y-k) = 1/26$. Therefore, $p_C(y|x) = p_k(y - x \mod 26) = 1/26$. Now the result follows from Baye's theorem. □

Consider a Cryptosystem with perfect secrecy. From Baye's theorem we get, $p_P(y|x) = p_P(y)$ $\forall x \in \mathcal{P}, y \in \mathcal{C}$. We can assume that $p_C(y) > 0$ $\forall y \in \mathcal{C}$ since any ciphertext y with $p_C(y) = 0$ can be removed. For any $x \in \mathcal{P}$, $p_C(y|x) = p_C(y) > 0$. Hence, $\forall y \in \mathcal{C}$, there exists a key k such that $e_k(x) = y$. Therefore, $|\mathcal{K}| \geq |\mathcal{C}|$. Also, $|\mathcal{C}| \geq |\mathcal{P}|$ since every encryption function is injective. Therefore, we have the following important observation.

$$\text{For Perfect Secrecy, we need to have } |\mathcal{K}| \geq |\mathcal{P}|. \quad (1)$$

Theorem 2 (Shannon [4]) *Suppose that* $|\mathcal{K}| = |\mathcal{P}| = |\mathcal{C}|$. *We have Perfect Secrecy if and only if each key is used with equal probability* $\frac{1}{|\mathcal{K}|}$ *and* $\forall x \in \mathcal{P}, y \in \mathcal{C}$ *there exists a unique key* k *such that* $e_k(x) = y$.

Proof: Suppose that there is Perfect Secrecy. As observed earlier, $\forall y \in \mathcal{C}$ there is a key k such that $e_k(x) = y$. Therefore, $|\mathcal{K}| \geq |\mathcal{C}|$. Fixing a $x \in \mathcal{P}$, $|\mathcal{C}| = |\{e_k(x) : k \in \mathcal{K}\}| \leq |\mathcal{K}|$. But $|\mathcal{C}| = |\mathcal{K}|$. Therefore, there cannot exist distinct keys k_1 and k_2 such that $e_{k_1}(x) = e_{k_2}(x) = y$.
Let $n = |\mathcal{K}|$ and $\mathcal{P} = \{x_i : 1 \leq i \leq n\}$. Fix a $y \in \mathcal{C}$ and name keys $k_1, \ldots, k_n$ such that $e_{k_i}(x_i) = y \ \forall i$. Using Baye's theorem and the fact that $p_P(x_i|y) = p_P(x_i) \ \forall i$ we have, $p_P(x_i|y) = \frac{p_C(y|x_i)p_P(x_i)}{p_C(y)} = p_P(x_i)$.
Hence $p_K(k_i) = p_C(y) \ \forall i$. In other words, all keys are used with equal probability which must be $1/|\mathcal{K}|$.
The proof of the converse is exactly similar to that of theorem 1. □

The Vernam Onetime Pad is a well known example of a Perfect Secrecy Cryptosystem. It is defined below.

The Vernam Onetime Pad

> Let $n \geq 1$ be an integer. $\mathcal{P} = \mathcal{K} = \mathcal{C} = (\mathcal{Z}_2)^n$, i.e. bit strings of length n. For any key $k \in \mathcal{K}$ we define $e_k(x) = d_k(x) =$ the vector sum modulo 2 (XOR) of k and x.

The Perfect Secrecy of this pad can be easily proved using theorem 2. Seeing this, one might wonder why this cannot be used for commercial applications. However, as each key is used for only one encryption and $|\mathcal{K}| \geq |\mathcal{P}|$ a new key needs to be sent across a secure channel each time!! Moreover, as many bits are required to represent the key as are needed to represent the plaintext. Therefore the secure channel used to send the key might as well be used to send the plaintext message too! Hence this method is not practical. This Vernam Onetime Pad was long thought to be unbreakable but this was first proved by Shannon in 1949 using theorem 2.

From the next section onwards, we will relax the condition that each key must be used for only one encryption, and analyze the chances of success of a ciphertext-only attack.

4 Entropy

In this section, we will deal with the notion of *entropy* which is borrowed from probability theory. This concept was first introduced by Shannon [3] in 1948.

Let a random variable X take a finite set of values corresponding to a probability distribution $p(X)$. The *entropy* of X, denoted as $H(X)$, is the information gained by an event taking place according to the distribution $p(X)$. In other words, if the event has not yet taken place, the entropy is a measure of the uncertainty about its outcome.

For example, in the toss of a coin, the probability of occurrence of a head and that of a tail are both $1/2$. The entropy is 1 bit, since we can encode the outcome with 1 bit. Similarly, for n tosses of a coin, the entropy is n.
Suppose that a random variable X take values x_1, x_2, x_3 with probabilities $1/2, 1/4$ and $1/4$ respectively. The most efficient encoding (0 , 10 , 11) requires an average of $3/2$ bits. In general, an event occurring with a probability 2^{-n} can be encoded as a bit string of length n, and one occurring with a probability p as a bit string of length $-\log_2 p$. Therefore, we can formally define entropy as follows.

Definition 2 *The* Entropy *of a random variable X with a probability distribution $p(X)$ is $H(X) = -\sum p_i \log_2 p_i$, where p_i denotes $p(X = x_i)$. If $p_i = 0$, then $\lim_{x \to 0} x \log x = 0$. Therefore we ignore such terms.*

We can see that if $p_i = 1/n \ \forall \ 1 \leq i \leq n$, then $H(X) = \log n$. Also $H(X) \geq 0$, with equality only when $p_i = 1$ for some i and $p_j = 0 \ \forall \ j \neq i$.

Given a random variable X, if $l(f)$ is the average length of an encoding f, then it can be shown that $H(X) \leq l(f_{\text{Huffman}}) < H(X) + 1$.

4.1 Properties Of Entropy

Definition 3 *A function is said to be* strictly concave *in an interval I if* $f(\frac{x+y}{2}) > \frac{f(x)+f(y)}{2} \quad \forall x, y \in I$.

We will use the following result from analysis extensively in the subsequent discussion.

Jensen's Inequality: Given a function f that is continuous and strictly concave in an interval I. If $\sum_1^n a_i = 1$ where $a_i > 0 \ \forall i$ then, $\sum_1^n a_i f(x_i) \leq f(\sum_1^n (a_i x_i))$ where $x_i \in I \ \forall i$. Equality occurs if and only if $x_1 = x_2 \cdots = x_n$.

Theorem 3 (Shannon [3]) *If X is a random variable with a probability distribution $p_1, p_2, \ldots, p_n$ such that $p_i > 0 \ \forall \ 1 \leq i \leq n$, then $H(X) \leq \log_2 n$ with equality occurring if and only if $p_i = 1/n \ \forall i$.*

Proof: We can easily see that $\log_2 x$ is strictly concave in $(0, \infty)$. Hence the expression for entropy can be bounded as follows

$H(X) = \sum_1^n p_i \log (1/p_i)$

$\leq \log \sum p_i (1/p_i)$ (From Jensen's Inequality)

$= \log n$. □

Theorem 4 (Shannon [3]) *If X and Y are random variables then $H(X,Y) \leq H(X) + H(Y)$ with equality occurring if and only if X and Y are independent.*

Proof: Let X be $\{x_i : 1 \leq i \leq n\}$ and Y be $\{y_i : 1 \leq i \leq n\}$. Let $p_i = p(X = x_i)$, $q_j = p(Y = y_j)$ and $r_{i,j} = p(X = x_i, Y = y_j)$ be the joint probability distribution. Obviously $p_i = \sum_{j=1}^n r_{i,j}$ and $q_j = \sum_{i=1}^m r_{i,j}$.

$H(X) + H(Y) = -\sum_i p_i \log p_i - \sum_j q_j \log q_j$

$= -(\sum_i \sum_j r_{i,j} \log p_i + \sum_j \sum_i r_{i,j} \log q_j)$

$= -\sum_i \sum_j r_{i,j} \log p_i q_j$.

Also, $H(X,Y) = -\sum_{i,j} r_{i,j} \log r_{i,j}$. Therefore,

$H(X,Y) - H(X) - H(Y) = \sum_i \sum_j r_{i,j} \log 1/r_{i,j} + \sum_i \sum_j r_{i,j} \log p_i q_j$
$= \sum_i \sum_j r_{i,j} \log \frac{p_i q_j}{r_{i,j}} \leq \log \sum_i \sum_j p_i q_j$
$= \log 1 = 0.$
(We could apply Jensen's Inequality here, since $r_{i,j}$ is a probability distribution).

Now consider the case of equality. There must exist a constant c such that $p_i q_j / r_{i,j} = c \ \forall i, j$. However, $\sum_i \sum_j r_{i,j} = 1$ and $\sum_i \sum_j p_i q_j = 1$.
Hence we have $c = 1$. In other words, $r_{i,j} = p_i q_j \ \forall i, j$ and therefore X and Y are independent. □

4.2 Conditional Entropy

Let X and Y be random variables. For a fixed value $y \in Y$, let $p(X|y)$ be the conditional probability distribution. Then we have,
$H(X|y) \ = \ -\sum_{x \in X} p(x|y) \log p(x|y)$.

Definition 4 *The* Conditional Entropy *of random variables X and Y denoted by $H(X|Y)$, is the weighted average of the entropies $H(X|y)$ and gives the average amount of information about X that Y reveals.*

$$H(X|Y) \ = \ -\sum_{y \in Y} \sum_{x \in X} p(y) p(x|y) \log p(x|y)$$

We can easily prove the following property of conditional entropy from theorem 4.

Theorem 5 (Shannon [3]) *If X and Y are random variables then $H(X,Y)$ $= \ H(Y) + H(X|Y)$*

As a direct corollary of theorems 4 and 5 we have,

Corollary 1 *For any two random variables X and Y, the conditional entropy $H(X|Y) \ \leq \ H(X)$ with equality occurring if and only if X and Y are independent.*

5 Spurious Keys & Unicity Distance

In this section, we shall consider the application of entropy to cryptography. By considering the conditional entropy $H(\mathcal{K}|\mathcal{C})$, which gives the information about the key distribution $\mathcal{K}$ revealed by the ciphertext, we will estimate the likelihood of success of Oscar's attack.

We call the conditional entropy $H(\mathcal{K}|\mathcal{C})$ as the *Key Equivocation* i.e. the information about the key revealed by the ciphertext.This conditional entropy can be expressed as follows.

Theorem 6 (Brassard [5]) *Let $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ be a Cryptosystem. Then,* $H(\mathcal{K}|\mathcal{C}) = H(\mathcal{K}) + H(\mathcal{P}) - H(\mathcal{C})$.

Proof: From theorem 5 we get $H(\mathcal{K}, \mathcal{P}, \mathcal{C}) = H(\mathcal{C}|\mathcal{K}, \mathcal{P}) + H(\mathcal{K}, \mathcal{P})$. Since a key along with a plaintext determines the ciphertext uniquely, $H(\mathcal{C}|\mathcal{K}, \mathcal{P}) = 0$.
Therefore $H(\mathcal{K}, \mathcal{P}, \mathcal{C}) = H(\mathcal{K}, \mathcal{P})$
$= H(\mathcal{K}) + H(\mathcal{P})$ (Because $\mathcal{K}$ and $\mathcal{P}$ are independent)
Similarly, $H(\mathcal{P}|(\mathcal{K}, \mathcal{C})) = 0$ giving $H(\mathcal{K}, \mathcal{P}, \mathcal{C}) = H(\mathcal{K}, \mathcal{C})$.
Thus $H(\mathcal{K}|\mathcal{C}) = H(\mathcal{K}, \mathcal{C}) - H(\mathcal{C}) = H(\mathcal{K}, \mathcal{P}, \mathcal{C}) - H(\mathcal{C})$
$= H(\mathcal{K}) + H(\mathcal{P}) - H(\mathcal{C})$. □

Consider a cryptanalyst Oscar trying to find the key, given a ciphertext. Oscar knows that the plaintext is English; therefore non-English yielding keys can be ruled out. For example, in the shift cipher, with *wnajw* as ciphertext, only *river* and *arena* can be possible plaintexts corresponding to keys 5 and 22. Therefore it is vital to know the number of keys which correspond to meaningful plaintexts; if only one such key exists then Oscar can always break the cryptosystem.

Definition 5 *A* Spurious Key *is one which is incorrect but corresponds to a meaningful plaintext.*

Let $\mathcal{P}^n$ be the random variable for n-grams of plaintext. We try to find a bound on the expected number of spurious keys.

For a natural language L the *entropy (per letter)* of L, denoted H_L, is a measure of the average information per letter in a "meaningful" plaintext. For a random sequence of alphabets, this entropy would be $\log_2 26 = 4.76$. As a first order approximation for $H(L)$, we can use $H(\mathcal{P})$ which is estimated to be 4.19 for English. This value is quite close to that for a random sequence of alphabets, which may appear surprising. However, successive letters in English are not independent. For example, a q is always followed by a u. Therefore, we can find the entropy of the probability distribution of digrams and divide by 2 to get better accuracy. Once again, for English, this entropy estimated from digrams turns out to be 1.95. Formally we have the following definitions.

Definition 6 *The* Entropy of a natural language *L is $H_L = \lim_{n\to\infty} \frac{H(\mathcal{P}^n)}{n}$. Here we account for all correlation between successive letters of the language by taking n-grams with $n \to \infty$.*

Definition 7 *The* Redundancy of a natural language *L is defined as $R_L = 1 - \frac{H_L}{\log_2 |\mathcal{P}|}$. Intuitively, we can see that $\log_2 |\mathcal{P}|$ is the amount of information each letter in a random text conveys whereas $\log_2 |\mathcal{P}| - H_L$ is the amount of information contained in each letter from L. Therefore R_L is the fraction of "excess" characters in the language.*

Statistics reveal that for English, $1 \leq H_L \leq 1.5$. Setting $H_L = 1.25$ we get a redundancy R_L of 75%!!! This of course does not mean that we can remove 75% of the letters in an English text and still hope to make sense, but that a careful encoding of the letters of the alphabet would reduce the size of the text by 75%.

Given the probability distributions on $\mathcal{P}^n$ and $\mathcal{K}$ we can derive the induced probability distribution on $\mathcal{C}^n$, the random variable for n-grams of ciphertext.

Fix a ciphertext $y \in \mathcal{C}^n$. We can see that the set of "possible" keys is $K(y) = \{k \in \mathcal{K} : \ni x \in \mathcal{P}^n,\ p_{P^n}(x) > 0 \text{ and } e_k(x) = y\}$. If y is the observed sequence of ciphertext, there are $|K(y)| - 1$ spurious keys. Therefore the

average number of spurious keys over all ciphertexts of length n is

$\bar{s_n} = \sum_{y \in \mathcal{C}^n} p(y)(|K(y)| - 1)$

$= \sum_{y \in \mathcal{C}^n} p(y)|K(y)| - 1.$

From theorem 6 we have $H(\mathcal{K}|\mathcal{C}^n) = H(\mathcal{K}) + H(\mathcal{P}^n) - H(\mathcal{C}^n)$. For large values of n, we can approximate $H(\mathcal{P}^n) \approx nH_L = n(1 - R_L)\log_2|\mathcal{P}|$. From theorem 3, we know that $H(\mathcal{C}^n) \leq n \log_2 |\mathcal{C}|$. Now, if we set $|\mathcal{C}| = |\mathcal{P}|$ then, $H(\mathcal{K}|\mathcal{C}^n) \geq H(\mathcal{K}) - n R_L \log_2 |\mathcal{P}|$.
Moreover, $H(\mathcal{K}|\mathcal{C}^n) = \sum_{y \in \mathcal{C}^n} p(y) H(\mathcal{K}|y)$
$\leq \sum_{y \in \mathcal{C}^n} p(y) \log_2 |K(y)| \leq \log_2 \sum_{y \in \mathcal{C}^n} p(y)|K(y)|$
$= \log_2(\bar{s_n} + 1)$.
Therefore $\log_2(\bar{s_n} + 1) \geq H(\mathcal{K}) - n R_L \log_2 |P|$. Hence we have the following theorem.

Theorem 7 (Brassard [5]) *If $|\mathcal{P}| = |\mathcal{C}|$ and keys are chosen equiprobably then, given a ciphertext of length n, for large n the expected number of spurious keys can be lower bounded by the following relation.*

$$\bar{s_n} \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1$$

This estimate is not accurate for small values of n since then the approximation $H_L \approx H(\mathcal{P}^n)/n$ will not be valid.

Definition 8 *The* Unicity Distance *of a Cryptosystem is that value of n at which $\bar{s_n} \to 0$. It is denoted by n_0. At the Unicity distance, the opponent can compute the key uniquely, given enough computing time. From theorem 7, we get the unicity distance as,*

$$n_0 \approx \frac{\log_2 |\mathcal{K}|}{R_L \log_2 |\mathcal{P}|}$$

For the substitution cipher with $\mathcal{P} = \mathcal{K} = \mathcal{C} = \mathcal{Z}_{26}$, we can estimate the unicity distance as given below.

$|\mathcal{P}| = 26$ and $|\mathcal{K}| = 26!$. Setting $R_L = 0.75$ as calculated earlier, we get $n_0 \approx \frac{88.4}{(0.75 \times 4.7)} = 25$.

6 Product Cryptosystems

In order to make the job of breaking the cryptosystem more difficult, we could use the product of two cryptosystems to encode the mesages. Here, we encrypt the given message first with one cryptosystem and then encrypt the resultant ciphertext using the next cryptosystem. We consider only *Endomorphic* cryptosystems i.e., those where $\mathcal{C} = \mathcal{P}$.

Given two cryptosystems $S_1 = (\mathcal{P}, \mathcal{P}, \mathcal{K}_1, \mathcal{E}_1, \mathcal{D}_1)$ and $S_2 = (\mathcal{P}, \mathcal{P}, \mathcal{K}_2, \mathcal{E}_2, \mathcal{D}_2)$, we define the product cryptosystem $S_1 \times S_2$ as $(\mathcal{P}, \mathcal{P}, \mathcal{K}_1 \times \mathcal{K}_2, \mathcal{E}, \mathcal{D})$.
The encryption and decryption functions are defined as $e_{(k_1,k_2)}(x) = e_{k_2}(e_{k_1}(x))$ and, $d_{(k_1,k_2)}(y) = d_{k_1}(d_{k_2}(y))$. The probability distribution of keys in the product cryptosystem is given by $p_K(k_1, k_2) = p_{K_1}(k_1)p_{K_2}(k_2)$.

The product operation on cryptosystems need not always be commutative, but is always associative. A cryptosystem S is said to be *idempotent* if $S^2 = S$. Many common ciphers like the Shift Cipher, the Affine Cipher and the Vignere Cipher are all idempotent. If a cryptosystem S is idempotent then there is no point in using S^2 to encrypt instead of S since each extra key is a waste. Otherwise, we could iterate the encryption process to use S^2 rather than S. For example, the *Data Encryption standard* uses 16 iterations.

If S_1 and S_2 are both idempotent and they commute, then $S_1 \times S_2$ is also idempotent (since $(S_1 \times S_2) \times (S_1 \times S_2) = S_1 \times (S_2 \times S_1) \times S_2 = (S_1 \times S_1) \times (S_2 \times S_2) = S_1 \times S_2$). Therefore to get a simple non-idempotent cryptosystem, we can simply take the product of two different cryptosystems which don't commute.

7 Conclusion

In this report, we have dealt with the major ideas in Shannon's theory of cryptography. Starting with the basic notion of Perfect Secrecy, we have looked at Shannon's treatment of Entropy. Then we applied this concept of Entropy to look at the likelihood of success of a ciphertext-only attack using

the idea of Spurious Keys. Finally, we concluded with a brief description of Product Cryptosystems. These concepts laid the first formal basis for cryptography and the notion of security of a cryptosystem upon which further theories were built.

References

[1] "Cryptography, Theory and Practice", Douglas R Stinson. CRC Press.

[2] "Applied Cryptography", Bruce Schneier. John Wiley & Sons.

[3] C F Shannon, "A mathematical theory of communication". *Bell Systems Technical Journal*, **27**(1949), pp. 379-423.

[4] C F Shannon, "Communication theory of secrecy systems". *Bell Systems Technical Journal*, **28**(1949), pp. 656-715.

[5] P Beauchemin and G Brassard, "A generalization of Helloni's extension to Shannon's approach to cryptography". *Journal of Cryptography*, **1**(1988), pp. 129-131.