VE475 Project1: Shannon's Theory and Cryptography

Yichen Hong, 518370910011
Kaibin Wang, 518021911059
Taoyue Xia, 518370910087
Wenjie Xianyu, 518370910133

June 2021

Abstract

In this project we look into "Shannon's Theory" and discuss its relation with cryptography. Firstly, we give an introduction of Claude Shannon and his theory of secrecy systems. Then we dig deep into "Shannon's Theory" and analyze several parts: perfect secrecy, entropy, equivocation, redundancy and unicity distance. Finally, we introduce some applications of Shannon's Theory on Cryptography, which involve IT-primitives, Hellman's extention and geeneralized random oracles.

Shannon's Theory has laid a foundation for information technology and cryptography in modern world, it is quite important and interesting. Moreover, cryptography is such a comprehensive subject with super magic, and the exploration of it will never stop.

Key words: Shannon's Theory, perfect secrecy, entropy, unicity distance, applications

Contents

1 Introduction

1.1 Claude Shannon

Claude Elwood Shannon was an American mathematician and cryptographer, who is known for creating the information theory. Shannon's most well-known paper is "A Mathematical Theory of Communication", which he published in 1948. In the paper, he introduced information entropy as a measure of the uncertainty reduced by the plaintext message.[8]

In 1949, he published another notable paper "Communication Theory of Secrecy Systems", in which he introduced some fundamental mathematical theories of cryptography.[8] "Perfect secrecy" is one of the most important theories.

1.2 Secrecy systems

In the paper "Communication Theory of Secrecy Systems", Shannon introduced three types of secrecy systems:

- **concealment systems**
- **privacy systems**
- **"true" secrecy systems**

Concealment systems contains methods such as using invisible ink, concealing messages in an innocent text, etc., which uses physical approaches to protect the real message from enemies.

Privacy systems contains methods such as speech inversion, which needs special equipment to recover the real message. It is a technical issue to attack the system.

"True" secrecy systems, which are the main concern of our project, uses ciphers and codes to conceal the real messages. As it is possible for enemies to intercept and record the transmitted ciphertext, we need to make the encryption algorithm as perfect as possible to prevent such occasion from happening. Below is the schematic of how this kind of systems works:

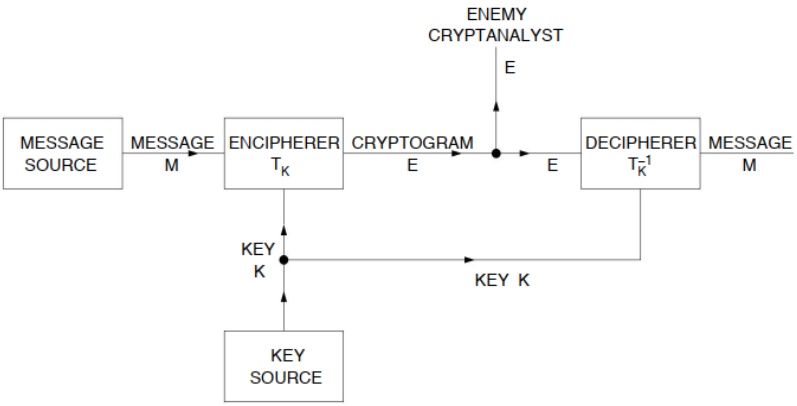

Figure 1: Schematic of a general secrecy system

2 Shannon's Theory

2.1 Perfect Secrecy

Suppose there is an adversary who is capable of learning the probability distribution of the message and the encryption scheme, as well as intercepting the ciphertext, but not the key used to encrypt the message, i.e., launching a ciphertext-only attack. Then, an encryption scheme is perfect secrecy when the adversary' s observing the ciphertext has no effect on his learning the plaintext.

The textbook Introduction to Modern Cryptography gives the definition of perfect secrecy as the following:

An encryption scheme (***Gen*, *Enc*, *Dec***) with message space $\mathcal{M}$ is perfectly secret if for every probability distribution for M, every message $m \in \mathcal{M}$, and every ciphertext $c \in \mathcal{C}$ for which $Pr[C = c] > 0$ [4]:

$$Pr[M = m \mid C = c] = Pr[M = m] \tag{2.1}$$

The formula above has an equivalent form. That is, for any message $m,\ m' \in M$, every cryptogram $c \in C$:

$$Pr[Enc_K(m) = c] = Pr[Enc_K(m') = c] \tag{2.2}$$

The textbook also provides a lemma:

> ***Lemma 1.*** An encryption scheme (***Gen*, *Enc*, *Dec***) with message space $\mathcal{M}$ is perfectly secret if and only if Eq. (2.2) holds for every $m,\ m' \in M$ and every $c \in C$ [4].

In plain words, ***Lemma 1*** says that for any $m,\ m' \in M$, the probability distributions of the ciphertext are the same, thus implying nothing about the plaintext.

2.2 Shannon's Theorem

The textbook defines Shannon' s Theorem as [4]:

Let (***Gen*, *Enc*, *Dec***) be an encryption scheme with message space M, for which $|M| = |K| = |C|$. The scheme is perfectly secret if and only if:

1. Every key $k \in K$ is chosen with (equal) probability $1/|K|$ by ***Gen***.
2. For every $m \in M$ and every $c \in C$, there is a unique key $k \in K$ such that $Enc_k(m)$ outputs c.

Since it is an "if and only if" relation, we first start with the proof of the sufficiency of the conditions, i.e., if the two conditions are satisfied, then the encryption scheme is perfectly secret.

Given condition 2, we have:

$$Pr[C = c|M = m] = Pr[Enc_k(m) = c] = 1/|K|$$

Therefore, given the property |M| = |K| = |C|, by Bayes' Theorem,

$$\begin{aligned} Pr[M = m|C = c] &= \frac{Pr[C = c|M = m] \cdot Pr[M = m]}{Pr[C = c]} \\ &= \frac{\frac{1}{|K|} \cdot \frac{1}{|M|}}{\frac{1}{|C|}} \\ &= \frac{1}{|M|} = Pr[M = m] \end{aligned}$$

According to the definition of perfect secrecy and Eq. (2.1), the encryption scheme is perfectly secret.

Next we deal with the necessity of the conditions, i.e., given that the encryption scheme is perfectly secret, then the two conditions should be satisfied.

For any $c \in C$, there exists m such that $Pr[Enc_K(m) = c] \neq 0$. Then, ***Lemma 1*** gives that for any $m,\ m' \in M$,

$$Pr[Enc_K(m) = c] = Pr[Enc_K(m') = c] \neq 0$$

Let $K_i \in K$ denote the nonempty set for m_i that $Enc_{K_i}(m_i) = c$. When $m_i \neq m_j$, K_i and K_j must be disjoint, otherwise the same key for different messages will give the same ciphertext. And since $|K| = |M|$, for any $m_i \in M$ we have:

$$|K_i| = 1$$

and there stands condition 2, the existence and uniqueness of the key which gives $Enc_k(m) = c$. Again based on ***Lemma 1***, we shall have:

$$Pr[K = k_i] = Pr[Enc_K(m_i) = c] = Pr[Enc_K(m_j) = c] = Pr[K = k_j]$$

for any $m_i,\ m_j \in M$, making "every key $k \in K$ is chosen with equal probability $1/|K|$" [4], as in condition 1. Q.E.D.

2.3 Entropy

Traditionally, the term "*entropy*" is firstly used in thermodynamics, describing the degree of chaos in a system. The greater the entropy is, the less possible it is to predict the system's condition.

Similarly, in order to determine how many choices are involved in the selection of events, or in other words, how uncertain we are of the final output, Shannon defines "*entropy*" in information theory as [1]:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log(P(x_i))$$

Here, X represnets for a discrete random variable with values $\{x_1, \cdots, x_n\}$, and $P(X)$ is the probability mass function of X.

In a secrecy system, we can use Shannon's Entropy to calculate the amount of information produced by a message, or the uncertainty to determine a specific message M:

$$H(M) = -\sum P(M) \log(P(M))$$

Also, there is an uncertainty of the choice of key K given by:

$$H(K) = -\sum P(K)\log(P(K))$$

In the perfect secrecy section above, the amount of information $H(M)$ contained in the message is at most $\log n$. Then the information contained by a message can be completely concealed from enemies if the key uncertainty $H(K)$ is at least $\log n$.

2.4 Equivocation

Known as the conditional entropy, the equivocation about two random variables X and Y is given by:

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

where $P(x,y)$ denotes the joint probability of X and Y.

In Shannon's theory, equivocation can be used as a theoretical secrecy index. In secrecy systems, there are two important equivocations of the message M and key K, denoted as $H(K|E)$ and $H(M|E)$, where E stands for the cryptogram [1]:

$$H(M|E) = -\sum_{M,E} P(M,E)\log(P(M|E))$$

$$H(K|E) = -\sum_{K,E} P(K,E)\log(P(K|E))$$

From the definition of information entropy, we can easily obtain that $H(X,Y) \leq H(X) + H(Y)$. While the message and key are independent with each other, we can prove that:

$$H(M,K) = H(M) + H(K)$$

Combine the above equations, we can obtain:

$$H(M,K) = H(E,K) = H(E) + H(K|E)$$

The former equality is because the obtained information of M and K is equal to that of E and K. The second equality holds in perfect secrecy, as $P(K) = P(K|E)$.

Then we can obtain an equivocation formula of key from the two above equations,

$$H(K|E) = H(M) + H(K) - H(E)$$

Thus we can find out that if the enemy only knows the cryptogram, or ciphertext, the uncertainty of key is quite large as they can get many possible keys, but only one of them is correct.

2.5 Redundancy

To get a general knowledge of redundancy, we first introduce the concept of entropy rate, or the relativity between letters.

Firstly, the absolute rate is defined as R, where

$$R = \log |M|$$

Here $|M|$ denotes the scale of message space of a secrecy system.

Then we define the entropy of each symbol as r, to see the relativity between a symbol and the symbols already known as:

$$r = \lim_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_n) = \lim_{n\to\infty} H(X_n | X_1, \cdots, X_{n-1})$$

Then the relative redundancy D can be defined as:

$$D = \frac{R - r}{R} = 1 - \frac{r}{\log |M|}$$

When every symbol has little relativity with the previous symbol, which means every symbol in M is independent with each other, then we can get,

$$\begin{aligned} r &= \lim_{n\to\infty} \frac{1}{n} (- \sum_{m\in|M|^n} P(m) \log m) \\ &= \lim_{n\to\infty} \frac{1}{n} (- \sum_{m\in|M|^n} \frac{1}{|M|^n} \log \frac{1}{|M|^n}) \\ &= \frac{n \log |M|}{n} = \log |M| \end{aligned}$$

$$\text{Then } D = 1 - 1 = 0$$

Thus we can know that every symbol in a message can carry the amount of information of value r, and the redundancy shows the needless ratio of information. The smaller the redundancy of a message is, the better it is as it is harder to decrypt.

2.6 Unicity Distance

Unicity distance is denoted as the length of an intercepted ciphertext, which is needed to break the cipher system by eliminating irrelated spurious keys in an attack [3]. The unicity distance U can be shown as:

$$U = \frac{H(K)}{D}$$

Here U represents for the actual unicity distance, $H(K)$ denotes the key uncertainty, or the entropy of the key space, and D stands for the redundancy of plaintext in bits.

Therefore, it is better if the unicity distance of a secrecy system is larger, as it means that an enemy needs more amount of ciphertext to decrypt the system. If the unicity distance approaches infinity, the system is called an ideal secret system [1].

For example, for a one time pad with size close to infinite, we can calculate for the entropy of the key space K unbounded as infinity, so $U \to \infty$, thus it is an ideal secret system. However, it is impossible to reach infinity, so the key size should be as large as possible.

3 Some Applications of Shannon's Theory on Cryptography

3.1 Information-theoretic Primitive

In this part, We are going to introduce some of the applications or further studies which are based on Shannon' s Theory on cryptography. In traditional complexity-theoretic cryptography, primitive as a generalization of cryptosystem is an important research body. In information-theoretic cryptography, *information-theoretic primitive* (IT-primitive) is the basic abstraction.

Definition 1. An stateless IT-primitive is a function which provides access for $n \geq 2$ players $P_1, ..., P_n$. Given inputs $\{X_i\}$, outputs $\{Y_i\}$ follows a known probability distribution $P_{Y_1,...,Y_n|X_1,...,X_n}$.

This abstraction could be generalized into the IT-primitive with internal state if the possibility distribution is correlated with previous input and output chain. This mechanism can be proved equal to traditional primitive system.
In cryptological setup, examples [5] could be provided based on the IT-primitive system

1. **Noisy Channel:** $P_{Y|X} = \epsilon, x \neq y$, $P_{Y|X} = 1-\epsilon, x = y$. In (γ, δ) noisy channel, $\epsilon \in [\gamma, \delta]$ is chosen by an opponent.
2. **Noisy Source:** No inputs, $P_{Y_A Y_B Y_E}$ is arbitrary known.
3. **Oblivious Transfer ($\binom{2}{1}-OT$):** $P_{Y_B|m_0 m_1 X}(m_0, x) = \gamma$, $P_{Y_B|m_0 m_1 X}(m_1, x) = 1-\gamma$. In OT, A knows gets no output; in Weak OT (WOT), A gets partial output from B's choice.
4. **Key Agreement:** No input, $P_{Y_A Y_B} = \epsilon, Y_A = Y_B$, $P_{Y_E} P_{Y_A Y_B} = P_{Y_E Y_A Y_B}$.
5. **Secure Message Transmission:** $P_{Y_B|X_A}(y_B, x_A) = 1, y_B = x_A$
6. **Authenticated but not Confidential Transmission:** $P_{Y_B Y_E|X_A}(y_B, y_E, x_A) = 1, y_B = y_E = x_A$
7. **Broadcast Channel:** $P_{Y_B,...Y_N|X_A}(y_B, ..., y_N, x_A) = 1, y_B = ... = y_N = x_A$

3.2 Reduction of IT-primitive

Based on the definition of the IT-primitive, we could generalize primitive reduction. A reduction problem could be defined as:

Definition 2. Given an available primitive or family of primitives $\{G_n\}$, in a specific tolerance of cheating, construct a primitive H.

Cheating could be categorized in three well-defined forms.

1. **Active Cheater(s):** deviate from the protocol.
2. **Passive Cheater(s):** follow the protocol, while keeping track of all the information in order to do violating operations.
3. **Fail-corrupted Cheater(s):** break down at any time and will stop executing the protocal.

We could model the cheaters as a neutral opponent that might corrupt some of the players.

Reduction for a primitive that exists in nature will result in unconditionally-secure protocols that is practically available. Reduction for a primitive that can be approximately implemented or can be implemented cryptographically will result in computationally-secure cryptographic protocols. [1] Examples of primitive reduction are given,

1. Under privacy amplification, reduce $\binom{2}{1} - OT^k$ to $\binom{2}{1} - OT$. [6]
2. Secret-key agreement by public discussion from noisy channels is the reduction of key agreement to a certain type of noisy source.
3. Secret sharing can be phrased as a reduction of a commitment primitive to the primitive of bi-literal secure communication channels if only passive cheating happens. [5]

3.3 Transfer Primitives

We first conceptualize $\binom{n}{1} - OT$ (OT) as:

Definition 3. $\binom{n}{1} - OT$ accepts an input X from A, and an input C from B. Output for B follows a random variable Y_C followed by a known distribution $P_{Y_1,\ldots,Y_n|X}$ while for any P_X, Y_C is independent on X and A gets no output about C.

Based on the definition, an (α, β)-transfer is an OT with $H(X) = \alpha$ and $H(Y_i) \leq \beta$, and we assume X has no irrelevant information about Y_i. A transfer hides γ bits if we have $H(X|Y_i) \leq \gamma$

3.4 Hellman's Extension of the Shannon Theory Approach to Cryptography

As the perfect secrecy we have talked about before requires the key has the length similar to the text, it' s hard to be widely used in real life. But this does not mean that Shannon' s theory is not practical.

Hellman[7] briefly introduced an extension of the Shannon Theory approach to cryptography in 1977 based on the unicity distance shown before. He showed the security of many kinds of ciphers, including the random one and proved it with Shannon' s theory of information. He also provided an idea for different languages to find the fittest cipher.

Definition 4. A random cipher is one in which, for each (C_0, K_0) pair, $g(C_0, K_0)$ has a uniform marginal distribution on all $2^{R_0 N}$ messages. However, the $g(C, K)$ are dependent in that, for any set S of cryptograms not including C_0, the distribution of $g(C_0, K_0)$ given $g(C, K_0)_{C \in S}$ is uniform over all messages not in $g(C, K_0)_{C \in S}$. There is no dependence between $g(C, K_0)$ and $g(C', K_0)$, for $K_0 \neq K_1$. [7]

For a language that has absolute rate

$$R_0 = log_2 L,$$

We have the unicity distance U, or noted as N_0 as

$$N_0 = \frac{H(K)}{D},$$

where

$$D = R_0 - R,$$

which is the redundancy of the language. Then for the text with length $N < N_0$, it is considered safe.

For English, we have:

$$H(K) = log_2 26! \approx 88.4 bits$$

$$R_0 = log_2 26 \approx 4.7 bits/character$$

$$R = 1.5 bits/character$$

$$D = 3.2 bits/character$$

So N_0 is calculated by:

$$N_0 \approx 28 characters$$

So it's clear that, the randomly chosen cipher isn' t good enough for a long English text. But if the D of a language equals or is close to 0, which means most of the permutations of the characters of this language are meaningful, then the simple randomly chosen code can be safe for texts with any length.

Then, in order to improve the security, we can try to apply some rules to the ciphertext in order to increase the value of unicity distance. For example, if we keep the frequency of occurrence of each characters for the ciphertext, which has 12.3 percent E' s, 9.6 percent T' s, etc. than D will be decreased from 3.2 to 2.67. And if we encrypt them with substitution on pairs or on words, the value of D will be decreased to 0.72.

It's also possible to have the cipher with redundancy $D = 0$, if the plaintext we have to encrypt is not based on a meaningful language, but a list of numbers, the plaintext and the cyphertext will both be meaningless, where D equals zero, and then the security can be ensured.

From Bellman' s word, Shannon' s Theory is not directly applicable to designing practical cryptographic systems, but we can use it to make improvement, or as a guidance when designing the cipher system. One thing he has derived for it is the importance to remove noise from the text

3.5 Generalized Random Oracles

A good example of application of Shannon's Theory is Random Oracle system. Maurer[6] gave a brief introduction of Generalized Random Oracles, the definition is shown below:

Definition 5. A generalized random oracle (GRO) is characterized by 1) a set of query operations, each of which takes as input an argument from a certain domain and outputs a corresponding value in a certain range, and 2) a random experiment for which each elementary event in the sample space is a complete set of answers to all possible queries, with some probability distribution over the sample space.[6]

Generally speaking, Random Oralce is a system that can generate a random-like key-value pair. This actually depends on a pseudo-random function that can distribute the possibility of the result uniformly over the output range. This is, for each meaningful text M_0, the output should be a random-like C_0, and for all C_0, they also have the same possibility to be decrypted to all possible meaningful texts.

So,we can consider the Random Oracles as a safe method if the PRF is not able to be broken for a limited number of operations, the number of operations depend on the calculation ability. If so, the attacker can only consider Random Oracle as random and this fits D=0 in Shannon' s theory, and which is perfectly secret.

4 Conclusion

The article starts from Shannon's theory of information, introducing the definitions about information theory, including entropy, equivocation, redundancy and unicity distance. We gave a brief proof of perfect secrecy, concluding Shannon's opinion about cryptosystem. Based on Shannon's theory, information-theory cryptography could be derived. In the passage, IT-primitive and it's examples are introduced. Based on the definition, reduction of IT-primitive and concepts of transfer primitives could be established.

Based on the theoretic definitions and derivations, applications are introduced. Hellman's Extension gives Shannon's theory extended applications and generalized random oracles as an important concept in information-theoretic cryptography is introduced.

While the theory of information and cryptography are two topics that are tightly connected, Shannon's theory would be quite valuable in the further study of cryptography. Right now, most application of Shannon's theory on cryptography is based on the concept of entropy. By increasing the entropy of a cipher system, we can also improve the security of it. Besides, we still have many parts of Shannon's theory that have not been applied into cryptography, which still have to potential to be developed.

References

[1] Shannon, Claude. *Communication Theory of Secrecy Systems*, Bell System Technical Journal, vol. 28(4), pp. 656–715, 1949.

[2] Shannon, Claude. *A Mathematical Theory of Communication*, Bell System Technical Journal, vol. 27(3), pp. 379–423, 1948.

[3] Alfred, Menezes, Paul, Oorschot, Scott, Vanstone. "Chapter 7 - Block Ciphers". *Handbook of Applied Cryptography*. p. 246.

[4] Katz, Jonathan, and Yehuda Lindell. *Introduction to Modern Cryptography*, 3rd ed., CRC Press, pp. 27-36, 2021.

[5] Maurer, Ueli. "Information-Theoretic Cryptography (Extended Abstract)." *Annual International Cryptology Conference*, 1999.

[6] Brassard, G., Crépeau, C. & Wolf, S. "Oblivious Transfers and Privacy Amplification". *J Cryptology* 16, 219–237, 2003. https://doi.org/10.1007/s00145-002-0146-4

[7] Hellman, M. E. . "An extension of the Shannon theory approach to cryptography." *IEEE Transactions on Information Theory* 23.3, pp. 289-294, 1977.

[8] James, Ioan (2009). "Claude Elwood Shannon 30 April 1916 –24 February 2001". *Biographical Memoirs of Fellows of the Royal Society*. pp: 257–265, 2009. doi:10.1098/rsbm.2009.0015