Information-theoretic Cryptography

Hermann Gruber, papro.soft GbR

June 6, 2005

Abstract

In 1949, Shannon published the paper "Communication theory of secrecy systems". This constituted a foundational treatment and analysis of encryption systems. He transferred the methods of information theory, originally developed as a mathematical model for communication over "noisy" channels to the setting of cryptosystems. We give a brief introduction into his most outstanding ideas, such as the notions of perfect/provable security, and statistical/information-theoretic concepts like entropy, key equivocation, and unicity distance.

Contents

1 Introduction

In 1949, Shannon published the paper "Communication theory of secrecy systems" [Sha49]. This constituted a foundational treatment and analysis of encryption systems. He transferred the methods of information theory, originally developed as a mathematical model for communication over "noisy" channels [Sha48], to the setting of cryptosystems. This includes the concepts of entropy and conditional entropy as measures for the secrecy of a system. Some of his seminal ideas inspired upcoming round-based encryption systems, such as the Data Encryption Standard [FIP93] or the modern Advanced Encryption Standard [FIP01]. But his most relevant achievement was his proof of the existence of systems with perfect secrecy. Informally stated, a system provides perfect secrecy if an intercepted cryptogram unveals no information about the plaintext, even provided the attacker has unbounded computational power. This is indeed a very strong assumption. The secrecy of most systems used today stems from computational hardness of code breaking: It is unlikely that an attacker with bounded computational power can break the code within reasonable time - unless he knows about some yet undiscovered clues, such as a fast factorization algorithm. (It should be mentioned that Shannon also included an informal discussion of this approach.) But why don't we use perfect secrecy systems today if we could? The main drawback of these systems is that the length of the key must equal the length of the plaintext message. And the key can be used only once (the most prominent among these carries the name "One-Time Pad"). Hence, we will have to transfer a large amount of data over a secure channel. Despite this apparent objection to practical applications, such systems are in use. For example, in a diplomatic context, two persons may agree at some time about a key, and both keep it in a secure place until it is needed. Then they can communicate in urgent cases over an insecure channel, while having achieved perfect secrecy. But most practical cryptosystems re-use a key more than once. Since in natural language there are symbols which appear more often than others, statistical attacks are possible. This is a principal weakness known to cryptanalists for a long time. For example, suppose an english message M is encrypted using a simple substitution (say every occurence of E is replaced by Q, and so on). For short ciphertexts, there are usually many reasonable plaintext "candidates" (and candidates for the key). Among these candidates, only one is correct; the others are spurious keys. With the length of the message, the probability for some few candidates increases, while it decreases for the others. Finally, there is a length where a single plaintext is most likely, whereas all others have probabilities near zero. This length is called unicity distance. Shannon formalized this idea and gave an estimate for this measure. The relevance of Shannon's pioneering work for modern systems is twofold: most today's cryptosystems do not incorporate perfect secrecy; statistical/information-theoretic methods are not the main tools of modern cryptanalysis. However, there are applications where perfect secrecy is needed; and many of the Shannon's original ideas are incorporated in the design of modern encryption algorithms. The paper is organized as follows:

After this introduction, we give a mathematical definition of what constitutes a cryptosystem in the next chapter. Then we introduce the reader into Shannon's term of "perfect secrecy systems", culminating in a complete characterization of these systems. Before we go over to the analysis of imperfect cryptosystems, we familiarize the reader with basic concepts of information theory. Finally, we end up with some concluding remarks on the impact of Shannon's work on today's developments in cryptography.

2 Cryptosystems

Before we start off, we need to clarify what we are talking about:

Definition 2.1 *"Cryptography is [...] the study of means of converting information from its normal, comprehensible form into an incomprehensible format, rendering it unreadable without*

secret knowledge – the art of encryption."
(definition taken from "WIKIPEDIA, the free encyclopedia", at http://en.wikipedia.org)

A crpytosystem is then a technical unit providing some mechanism for converting normal messages into incomprehensible form and back. This definition, although concise and clear, is too fuzzy to allow a rigorous mathematical analysis.

So, in the literature, a cryptosystem is usually defined as a protocol between two partners, say Alfons and Boris:

- Alfons and Boris agree about a secret key k, and the key is transmitted over a secure channel.
- Alfons encrypts $T_k(m) = c$, sends it to Boris
- Boris decrypts $T_k^{-1}(c) = m$
- Ivan (the Terrible) intercepts c and tries to figure out m

Ivan will of course also try to find out the key to be able to decrypt future messages – which would be the worst case scenario for Alfons and Boris.

It is common use to assume that the encryption mechanism is publicly known; that is the secrecy is based only on secrecy of the key. This is known as Kerckhoffs' principle [Ker83].

Some specific attacks are considered here: Assume Ivan has passed by Alfons' rear window, while he typed a plaintext message he lateron encrypted and sent to Boris. Alfons also intercepted the corresponding ciphertext. He now wants to obtain the key within reasonable (parallel) computing time. Such a scenario is called known-plaintext attack.

Other scenarios consider cases where a certain amount of plaintext can be chosen, and even where the attacker can adaptively choose a new plaintext depending on the last encryption result. All these scenarios assume an attacker with bounded resources.

It seems very hard to prove that there is no algorithm running within N steps which can be used to break a cryptosystem. Instead, designers of cryptosystems try to avoid weakness against known attacks.

Shannon's point of view is slightly different: His model is biased on the standpoint of a possible attacker who wants to gain (statistical) information about the plaintext messages, or the key. He assumes that the attacker has *unbounded* resources, but can by *no means* obtain a plaintext-ciphertext pair. In this setting, Alfons is considered as a statistical information source (a random variable), of which Ivan knows some a *priori probability distribution.* For example, the message

SENDSUPPLIESASAP

is much more likely than the following rather dull message

DXKLUNQUEOEFLACD

especially if Ivan knows *a priori* that Alfons is running out of beer. Shannon also assumes that Ivan knows about Alfons' habit of key choice, but using todays' electronic cryptosystems we mostly assume that the keys are chosen equiprobably.

These two ingredients take responsibility for the produced ciphertext, so it seems reasonable to ask to what extent the information is conveyed by the ciphertext.

3 Perfect Secrecy Systems

3.1 Definition of Perfect Secrecy

More formally, given two random variables M and K representing the finite set of possible plaintext messages and of possible keys, each with an associated a priori probability distri-

bution, how much differs the a posteriori probability distribution when Ivan has observed a certain ciphertext c?

A secrecy system is called perfectly secure if the a priori probability distribution equals the a posteriori distribution.

Definition 3.1 (Shannon 1949) *A cryptosystem with probability distributions on message space M and keyspace K is said to be perfectly secret, if for* all *ciphertext messages c and* all *messages m holds*

$$Pr[m \mid c] = Pr[m]$$

An example for a perfect secrecy system is the following:

Example 3.1 ("One-Time Pad" (Gilbert Vernam 1926, patented 1919))

- $M = K = C = \{0,1\}^n$
- *keys are chosen equiprobable*
- *encryption/decryption: bitwise modulo-2-addition of key and message*
- *key is only used* once.

(Note that the key is as long as the message.) Can we *prove* that this system is perfectly secure?

3.2 Development of the Tools

Suppose Ivan intercepts a ciphertext c, and wants to obtain a posteriori (conditional) distribution on M, depending on c. A very common and useful theorem for dealing with conditional probabilities is Bayes' theorem:

Theorem 1 (Bayes' theorem) *Let C and M denote random variables. If $P[C = c] > 0$, then*

$$P[M = m \mid C = c] = \frac{P[C = c \mid M = m]P[M = m]}{P[C = c]}$$

Furthermore, C and M are independent iff $P[M = m \mid C = c] = P[M = m]$.

Next, we define the set of possible keys for a given cipher c:

$$K(c) = \{k \in K \mid \exists m \in M : T_k(m) = c\}$$

And, accordingly the set of possible keys for a given message m and a given ciphertext c is defined as

$$K(c, m) = \{k \in K \mid T_k(m) = c\}.$$

Then, we have obviously

$$P[C = c] = \sum_{k \in K(c)} P[K = k]P[M = T_k^{-1}(c)]$$

and

$$P[C = c \mid M = m] = \sum_{k \in K(c,m)} P[K = k].$$

Putting these into Bayes' formula yields

$$\begin{aligned} P[M = m \mid C = c] &= \frac{P[C = c \mid M = m]P[M = m]}{P[C = c]} \\ &= \frac{\sum_{k \in K(c,m)} P[K = k]P[M = m]}{\sum_{k \in K(c)} P[K = k]P[M = T_k^{-1}(c)]} \end{aligned}$$

In this way, we can *always* compute the a posteriori probability distribution from the known a priori probability distribution and an intercepted ciphertext; Furthermore, this calculation can be carried out without much effort.

3.3 One-time Pad provides Perfect Secrecy

Now we are ready to prove that the One-time Pad does indeed provide perfect secrecy.

Theorem 2 *The One-time Pad is a perfect secrecy system.*

Proof We have to show that for every possible messages m and every possible ciphertext c holds $P[M = m \mid C = c] = P[M = m]$.

For every k, we have $P[K = k] = \left(\frac{1}{2}\right)^n$, as the keys are chosen equiprobably. So we obtain

$$P[C = c] = \left(\frac{1}{2}\right)^n \sum_{k \in K(c)} P[M = T_k^{-1}(c)]$$

for every possible ciphertext c.

Next, see that for every message-ciphertext pair $\langle m, c\rangle$, there is a *unique* key that can be used for m to obtain c; In other words, $|K(c, m)| = 1$ for every pair $\langle m, c\rangle$ Using this knowledge in the above equation, we get

$$\sum_{k \in K(c)} P[M = T_k^{-1}(c)] = \sum_{m \in M} P[M = m] = 1;$$

and we conclude

$$P[C = c] = \left(\frac{1}{2}\right)^n.$$

That is, C is distributed equiprobably over the ciphertext space.

We had seen in the previous section that $P[C = c \mid M = m] = \sum_{k \in K(c,m)} P[K = k]$. Now, for every possible message-ciphertext pair $|K(c, m)| = 1$, and the following holds:

$$P[C = c \mid M = m] = \sum_{k \in K(c,m)} P[K = k] = \left(\frac{1}{2}\right)^n$$

Surely, this is the "wrong" probability distribution, as we actually want to determine $P[C = c \mid M = m]$. But we can use Bayes' formula to "flip" the variables in a conditional distribution:

$$\begin{aligned} P[M = m \mid C = c] &= \frac{P[C = c \mid M = m]P[M = m]}{P[C = c]} \\ &= \frac{P[C = c \mid M = m]P[M = m]}{\left(\frac{1}{2}\right)^n} \\ &= \frac{(1/2)^n \, P[M = m]}{(1/2)^n} \\ &= P[M = m]. \end{aligned}$$

□

So, we have *proved* that there exist examples of perfect secrecy systems, and we have seen that One-time Pad is indeed a living example of this species.

3.4 Characterization of Perfect Secrecy

The main ideas from the previous chapter can be generalized to obtain sufficient conditions for perfect secrecy. Then it is not hard to show that all of these conditions are also necessary, and with not to much effort one obtains the following

Theorem 3 (Perfect Secrecy Theorem [Sha49]) *A cryptosystem provides perfect secrecy if and only if*

- $|M| = |C| = |K|$
- *every key is used with equal probability* $1/|K|$,
- *and for every message-cipher-pair* $\langle m, c\rangle$*, there is a unique key* k *with* $c = T_k(m)$.

(The proof can be found in [Sti02].)

Thus it can be seen that all perfect secrecy systems are very similar in nature – one might even be tempted that they are all just different "encodings" of the One-time Pad.

Note that perfect secrecy has a high price: From the above theorem, we can see that Perfect Secrecy is often impractical, as the key needs to be as long as the plaintext message – this can be seen from the fact that the message space has to be as large as the key space, and since all keys are chosen equiprobably, we cannot use compression to reduce the length of the key.

3.5 Conclusions about Perfect Secrecy

Some possible ways around are proposed in the following: We may make use of pseudo-random generators instead of true random choice for the One-Time Pad. This allows to dramatically reduce the size of the key which has to be transmitted. On the other hand, this provides no longer perfect secrecy.

We may rethink our model of secrecy: Is it too pessimistic to think of the enemy as having unbounded computational power? These considerations led to the development of other models for secrecy, where the enemy only has limited computational power, but may have other possibilities, for example he may have discovered a plaintext-ciphertext pair. Unfortunately, a mathematical analysis of this setting seems to be very hard – to the moment, no positive results about the secrecy of systems are proved without using additional assumptions.

Nevertheless, the One-Time pad is still in use for very critical purposes, such as in diplomatic or military contexts.

4 Information Theoretic Concepts

4.1 Entropy

The difficulties arising with perfect secrecy systems, namely the need for transmission of a large key over a secure channel, raises the following question: What if we use the same key more than once?

The mathematical analysis is again due to Shannon. He used a concept from information theory, namely *entropy*. Informally spoken, the entropy measures the *average* degree of *uncertainty* of a statistical quantity. Shannon gave the following definition along with a mathematical motivation for using entropy as a measure for information.

Definition 4.1 (Entropy, [Sha48]) *Let X be a random variable taking on the values 1,. . . ,n. Then*

$$H(X) = -\sum_{i=1}^{n} P[X = i] log_2 P[X = i]$$

Example 4.1 *If we throw a fair (Laplace) die, the result is an equiprobably distributed random variable X. Its entropy can be calculated as*

$$H(X) = -\sum_{i=1}^{6} \frac{1}{6} log_2 \frac{1}{6} = log_2 6$$

In general, the entropy of a random variable taking on values from 1 *to* n*, the entropy is bounded above by the one of an equiprobable distribution:*
X *ranges from* 1 *to* $n \Rightarrow H(X) \leq \log_2 n$.

4.2 Conditional Entropy

In the theory of probability, we often consider conditional random variables. The following generalization of the definition is – besides of its importance for information theory – of particular use for the analysis of secrecy systems:

Definition 4.2 (Conditional Entropy, [Sha48]) *Let Y be another random variables, taking values 1,...,m The conditional entropy $H(X \mid Y)$ is*

$$H(X \mid Y) = \sum_{j=1}^{m} p(Y = j) H(X \mid Y = j)$$

We can state that conditional entropy measures the average uncertainity about X *given observations of the variable Y.* In the context of (imperfect) secrecy systems, we can use this quantity to answer the following natural question: How much average uncertainty remains about the key remains provided we have intercepted the ciphertext?

For a secrecy system, we call the conditional entropy $H(K \mid C)$ the *key equivocation.*

Shannon found that for this quantity holds [Sha49]

Theorem 4

$$H(K \mid C) = H(M) + H(K) - H(C)$$

(The proof can be found in [Sti02],Thm. 2.10)

As an immediate implication of the above theorem, we have $H(K \mid C) = H(K)$ in the case of a perfect secrecy system, and this is a necessary and sufficient condition for perfect secrecy. That is, uncertainty about the key does not decrease with knowledge of the ciphertext.

5 Statistical Analysis of Imperfect Cryptosystems

5.1 Entropy of Natural Language

If a cryptosystem is not perfectly secure, we can deduce that some keys are more likely than others – provided we encode some information with a "reasonable" probability distribution. In particular, we analyze the situation where the same key is reused over and over again to encode a sequence of messages.

Suppose the same key is used over and over again to encode a long message, and Ivan knows "a priori" that Alfons and Boris exchange messages in natural language. So, Boris can reduce the message space to a smaller space of "likely messages" – and hence the key space.

To formalize our intuition, we need a notion of the entropy of a natural language, such as English. Suppose we have intercepted a part of a ciphertext having three (n) letters. We know that some "tri-grams" (n-grams) are very common in English, such as 'THE' is much more likely to occur than 'SRX' (although one should be aware that the situation is entirely different in marketing english: The company Linksys® uses "SRX" as acronym for "Speed and Range eXpansion" [oCSI]; In general, words on three letters with an "X" are very common in technical marketing speech!)

We define M^n as the probability distribution for plaintext messages of length n, and correspondingly C^n for the ciphertexts. Note that C^n is far from being equiprobably distributed, as we reuse the same key n times!

The entropy of the english language is then defined as

$$H_L = \lim_{n\to\infty} H(M^n)$$

An immediate upper bound is the entropy of random strings over the 26-letter alphabet, that is $H_L \leq \log_2 26 \approx 4.7$; A better estimate is obtained by sweeping large amounts of english texts. Such experiments led to the empirical result $1.0 \leq H_L \leq 1.5$.

5.2 Unicity Distance.

We go on with the assumption that the same key was used n times, and a ciphertext of length was intercepted. In this situation, a cryptanalyst can deduce that a small subset of the keyspace contains the key which was actually used for encrypting. Of course, only *one* key was actually used; this subset of the keyspace – excluding the correct key – is called a set of *spurious keys*.

More formally, $S_n(c)$ denotes the set of keys k for which, given a ciphertext c of length n, there is an decryption $T_k^{-1}(c) = m$ such that m is a *meaningful* english text. (To avoid complications, we omit a detailed definition of the term "meaningful"...)

The longer the intercepted ciphertext message, the more keys can be ruled out, as they do not encode "sensible" messages; and the smaller is the remaining set of spurious keys. We are interested to what extent the number of spurious keys reduces, if we intercept more and more ciphertext, and in particular we want to determine the amount of ciphertext where the number of spurious keys approaches zero. This amount is called the *unicity distance*.

The average number of spurious keys over all possible ciphertexts of length n is denoted by $\bar{s}_n$, and it holds

$$\bar{s}_n = \sum_{c\in C^n} Pr[c](|S_n(c)| - 1) = \sum_{c\in C^n} (Pr[c]S_n(c)) - 1$$

We want to calculate the value $\bar{s}_n$, and in particular we are interested in the point n where the number of spurious keys approaches zero. We continue with some estimates; first we show how $\bar{s}_n$ can be related to the key equivocation of a ciphertext of length n:

$$\begin{aligned} H(K \mid C^n) &= \sum_{c\in C^n} Pr[c]H(K \mid c) \\ &\leq \sum_{c\in C^n} Pr[c]\log_2 |S_n(c)|, \end{aligned}$$

since we can assume $K \mid c$ ranges only over values encoding meaningful messages. The last term in the above inequality almost looks like our formula for the average number of spurious keys. From theory on convex functions, we can apply Jensen's inequality [Jen06] to obtain

$$\sum_{c\in C^n} Pr[c]\log_2(|S_n(c)|) \quad \leq \log_2\left(\textstyle\sum_{c\in C^n} Pr[c]|S_n(c)|\right) \quad = \log_2(\bar{s}_n + 1)$$

Recall from Theorem 4 that $H(K \mid C^n) = H(K) + H(M^n) - H(C^n)$. We can estimate that $H(M^n) \approx nH_L$ and $H(C^n) \leq n\log_2|C|$. Putting all things together, we get the following

Theorem 5 ([Sha49], see also [SL]) *For a cryptosystem in which keys are chosen equiprobably and $|M| = |C|$, we have the heuristic*

$$\bar{s}_n \geq |K|/(|M|)^{n(1-H_L)} - 1$$

To estimate the unicity distance, we simply put $s_n = 0$ in the above inequality. In particular, for modern block ciphers, we have a unicity distance which is lower than 2. This means that, by information-theoretic means, intercepting only two ciphertext blocks will be enough to reproduce the key. Fortunately, today's human or software attackers only have bounded computational power ...

6 Conclusion

Shannon's pioneering work is regarded by much people working in the field as *the foundation* of cryptography. We have looked at some of the most aspects of his work which count to the opinion of the author among the most relevant. This included the development of a model of a cryptosystem, with Alfons and the key generator as statistical information source. The enciphered text is transmitted over a public channel, while nobody but Alfons and Bob know anything about the secret key, and the plaintext - inlcuding a possible attacker Ivan. The attacker is assumed to have unbounded computational power, and – by the Kerckhoffs' principle – he knows the interior mechanism of the cryptosystem.

This principle of transparency is very important for most of modern cryptography, as the designer of a cryptosystem has to take care that secrecy is based *only* on the secret key being unknown. Today's cryptographic community mostly shows an offensive handling with this principle: All details of the design of modern cryptographic algorithms are publicly available. This allows the community to carry out a rigorous process of scrutiny. Potential flaws or weaknesses can be discussed publicly, and anyone not trusting the secrecy of the system is free to rule out his doubts himself.

Using concepts from probability and information theory, we have analyzed an example of a perfect secrecy system, that is, a system where an attacker cannot gain *any* kind of information from observing the ciphertext, regardless of his computational resources (in bth time and space) nor the length of the observed ciphertext. This example is the One-Time Pad, which was developed by the engineer Gilbert Vernam. We also stated a characterization of perfect secrecy systems, from which we saw that all such systems are essentially alike.

Perfect secrecy requires that the amount of key information to be transported over a secure channel is the same as the amount of information which we desire to encrypt. This is a disadvantage of perfect secrecy, and explains its little importance in practice. We used concepts from information theory to analyze the case where we cannot have perfect secrecy, but use the same key over and over again.

Using concepts from information theory, we saw that there exists a certain amount of ciphertext which is enough to uniquely determine not only the plaintext message, but the key which was actually used; and we can roughly estimate the size of this required amount.

Information theory can sometimes be a powerful tool to prove lower bounds on computational hardness – in general a very challenging task. We have shown that no algorithm can break a perfect secrecy system. Another example is the folklore result that any general (comparison-based,deterministic) algorithm sorting n items requires time $\Omega(n \log n)$. But, unfortunately, in most situations lower bounds from information theory perform "arbitrarily bad" – this is where we require additional (unproved) assumptions on which mathematical proofs of secrecy properties are based.

References

[FIP93] Federal Information Processing Standards. *Announcing the Standard for Data Encryption Standard (DES)*, 1993.

[FIP01] Federal Information Processing Standards. *Announcing Advanced Encryption Standard (AES)*, 2001.

[Jen06] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.*, 30:175–193, 1906.

[Ker83] A. Kerckhoffs. La cryptographie militaire i. *Journal des sciences militaires*, IX:5–138, January 1883.

[oCSI] Linksys® A Division of Cisco Systems Inc.

[Sha48] C.E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27(3):379–423, July 1948.

[Sha49] C.E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, October 1949.

[SL] Nigel Smart and John Malone Lee. Introduction to cryptography. Lecture notes of course COMS30124 at the University of Bristol.

[Sti02] D. R. Stinson. *Cryptography – Theory and Practice*, chapter Shannon's Theory. Discrete Mathematics and its Applications. Chapman and Hall / CRC, 2002.