Theoretical Security

Klaus Pommerening
Fachbereich Mathematik
der Johannes-Gutenberg-Universität
Saarstraße 21
D-55099 Mainz

February 6, 2000—English version July 30, 2014—last change
January 19, 2021

The theory of this section goes back to Claude Shannon[15] (with later simplifications by Hellman[8]). In his paper Shannon developed the first general mathematical model of cryptology as well as the analysis of cryptosystems by information theoretical methods. The basic question this theory asks is:

> *How much information about the plaintext is preserved in the ciphertext?*

(no matter how difficult or expensive the extraction of this information is.) If this information doesn't suffice to determine the plaintext, then the cipher is secure.

Shannon's ideas are based on the information theory that he had developed before [14].

The practical value of Shannon's theory is limited. But besides it there are almost no sufficient criteria for the security of cryptographic methods that are mathematically proved. In contrast there are lots of necessary criteria derived from cryptanalytic procedures. Lacking better ideas one tries to optimize the cryptographic procedures for these necessary conditions. We saw and shall see many instances of this in these lecture notes.

1 A Priori and A Posteriori Probabilities

Model Scenario

Consider

- a finite set $M_0 \subseteq M$ of possible plaintexts—for example all plaintexts of length r or of length $\leq r$,
- a finite set K of keys,
- a cipher $F = (f_k)_{k \in K}$ with $f_k \colon M \longrightarrow \Sigma^*$.

The restriction to a finite set M_0 allows us to handle probabilities in the naive way. It is no real restriction since plaintexts of lengths $> 10^{100}$ are extremely unlikely in this universe that has at most 10^{80} elementary particles.

Motivating Example

For English plaintexts of length 5 we potentially know exact a priori probabilities, say from a lot of countings. A small excerpt from the list is

Plaintext	Probability
hello	$p > 0$
fruit	$q > 0$
xykph	0
...	...

Now assume we see a monoalphabetically encrypted English text `XTJJA`. Without knowing the key—that is in a situation where all keys have the same probability—and without further context information we nevertheless assign to the single plaintexts different "a posteriori probabilities":

Plaintext	Probability
hello	$p_1 >> p$
fruit	0
xykph	0
...	...

Thus knowledge of the ciphertext alone (and knowledge of the encryption method) changed our information on the plaintext.

A "Bayesian" approach gives a general model of this observation.

Model

The probability of plaintexts is given as a function

$$P \colon M_0 \longrightarrow [0,1] \quad \text{where} \quad P(a) > 0 \quad \text{for all } a \in M_0 \quad \text{and} \quad \sum_{a \in M_0} P(a) = 1.$$

(This is the a priori probability of plaintexts.)

The probability of keys is likewise given as a function

$$P\colon K \longrightarrow [0,1] \quad \text{such that} \quad \sum_{k\in K} P(k) = 1.$$

(By abuse of notation denoted by the same letter P.) In general we assume a uniform distribution $P(k) = 1/\#K$ for all $k \in K$.

The probability of ciphertexts derives from the probabilities of plaintexts and keys, implicitly assumed as independently chosen:

$$P\colon \Sigma^* \longrightarrow [0,1], \quad P(c) := \sum_{a\in M_0} \sum_{k\in K_{ac}} P(a)\cdot P(k),$$

where $K_{ac} := \{k \in K \mid f_k(a) = c\}$ is the set of all keys that transform a to c.

Remark 1 Only finitely many $c \in \Sigma^*$ have $P(c) \neq 0$. These form the set

$$C_0 := \{c \in \Sigma^* \mid P(c) > 0\}$$

of "possible ciphertexts".

Remark 2 We have

$$\begin{aligned}
\sum_{c\in\Sigma^*} P(c) &= \sum_{c\in\Sigma^*} \sum_{a\in M_0} \sum_{k\in K_{ac}} P(a)\cdot P(k) \\
&= \sum_{a\in M_0} \sum_{k\in K} P(a)\cdot P(k) \\
&= \sum_{a\in M_0} P(a) \cdot \sum_{k\in K} P(k) \\
&= 1.
\end{aligned}$$

The conditional probability for a ciphertext to stem from a given plaintext $a \in M_0$ is modeled by the function

$$P(\bullet|a)\colon \Sigma^* \longrightarrow [0,1], \quad P(c|a) := \sum_{k\in K_{ac}} P(k).$$

Remark 3 $\sum_{c\in\Sigma^*} P(c|a) = \sum_{k\in K} P(k) = 1$.

Remark 4 $P(c) = \sum_{a\in M_0} P(a)\cdot P(c|a)$.

A Posteriori Probabilities of Plaintexts

The cryptanalyst is interested in the converse, the conditional probability $P(a|c)$ of a plaintext $a \in M_0$ if the ciphertext $c \in \Sigma^*$ is given.

First we describe the probability of the simultaneous occurrence of a and c as

$$P\colon M_0 \times \Sigma^* \longrightarrow [0,1], \quad P(a,c) := P(a) \cdot P(c|a).$$

Remark 5 Then

$$\sum_{a \in M_0} P(a,c) = \sum_{a \in M_0} P(a) \cdot P(c|a) = P(c).$$

The conditional probability of a plaintext is given by a function $P(\bullet|c)$ with $P(a,c) = P(c) \cdot P(a|c)$ by the Bayesian formula

$$P(a|c) := \begin{cases} \frac{P(a) \cdot P(c|a)}{P(c)} & \text{if } P(c) \neq 0, \\ 0 & \text{if } P(c) = 0. \end{cases}$$

Remark 6 $\sum_{c \in \Sigma^*} P(c) \cdot P(a|c) = \sum_{c \in \Sigma^*} P(a) \cdot P(c|a) = P(a)$ by Remark 3.

2 Perfect Security

Definition 1 The cipher F is called **perfectly secure** on M_0 (the finite set of all possible plaintexts) if $P(\bullet, c) = P$ on M_0 for all ciphertexts $c \in \Sigma^*$ of positive probability $P(c) > 0$.

Interpretation: This condition assures that the a posteriori probability $P(a|c)$ of each plaintext $a \in M_0$ is the same as the a priori probability $P(a)$. Or in other words, the cryptanalyst doesn't get any additional information on the plaintext by knowing the ciphertext.

Lemma 1 $\#M_0 \leq \#C_0$.

Proof. Let $l \in K$ be a fixed key with $P(l) > 0$. For every ciphertext $c \in f_l(M_0)$, say $c = f_l(b)$, we then have

$$P(c) = \sum_{a \in M_0} P(a) \cdot \sum_{k \in K_{ac}} P(k) \geq P(b) \cdot P(l) > 0.$$

Hence $c \in C_0$. From this follows that $f_l(M_0) \subseteq C_0$. Since f_l is injective also $\#M_0 \leq \#C_0$. $\diamond$

Lemma 2 *If F is perfectly secure, then $K_{ac} \neq \emptyset$ for all $a \in M_0$ and all $c \in C_0$.*

Proof. Assume $K_{ac} = \emptyset$. Then

$$P(c|a) = \sum_{k \in K_{ac}} P(k) = 0.$$

Hence $P(a|c) = 0 \neq P(a)$, contradiction. $\diamond$

Therefore each possible plaintext can be transformed into each possible ciphertext. The next lemma says that the number of keys must be *very* large.

Lemma 3 *If F is perfectly secure, then $\#K \geq \#C_0$.*

Proof. Since $\sum P(a) = 1$, we must have $M_0 \neq \emptyset$. Let $a \in M_0$. Assume $\#K < \#C_0$. Then there exists a $c \in C_0$ with $f_k(a) \neq c$ for every key $k \in K$, whence $K_{ac} = \emptyset$, contradiction. $\diamond$

Theorem 1 [Shannon] *Let F be perfectly secure. Then*

$$\#K \geq \#M_0.$$

That is the number of keys is at least as large as the number of possible plaintexts.

Proof. This follows immediately from Lemmas 1 and 3. ◇

Theorem 2 [SHANNON] *Let F be a cipher with*

$$P(k) = \frac{1}{\#K} \quad \textit{for all } k \in K$$

(that is all keys have the same probability) and

$$\#K_{ac} = s \quad \textit{for all } a \in M_0 \textit{ and all } c \in C_0.$$

with a fixed $s \geq 1$. Then F is perfectly secure. Furthermore $\#K = s \cdot \#C_0$.

Proof. Let $c \in C_0$ be a possible cipherext. Then for any possible plaintext $a \in M_0$:

$$\begin{aligned} P(c|a) &= \sum_{k \in K_{ac}} \frac{1}{\#K} = \frac{\#K_{ac}}{\#K} = \frac{s}{\#K}, \\ P(c) &= \sum_{a \in M_0} P(a) \cdot P(c|a) = \frac{s}{\#K} \cdot \sum_{a \in M_0} P(a) = \frac{s}{\#K} = P(c|a), \\ P(a|c) &= \frac{P(c|a)}{P(c)} \cdot P(a) = P(a). \end{aligned}$$

Therefore F is perfectly secure. The second statement follows from

$$K = \dot{\bigcup_{c \in C_0}} K_{ac}$$

for all $a \in M_0$. ◇

3 Examples of Perfect Security

Trivial Examples

Example 0: $\#M_0 = 1$. This example is cryptological nonsense since the cryptanalyst knows the only possible plaintext a priori. Hence she cannot gain any additional information on the plaintext by knowing the ciphertext.

Let $M_0 = \{a\}$. For all $c \in C_0$ trivially $P(a|c) = 1 = P(a)$. Hence F is perfectly secure, no matter how it is defined.

Example 1: $\#M_0 = 2$. The smallest nontrivial example involves two possible plaintexts. Without restriction we may assume that $M_0 = \{0, 1\} = C_0 = K$. Let f_0 be the identity map on $\{0, 1\}$, and f_1, the transposition of 0 and 1. Furthermore let the two keys 0 and 1 have the same probability: $P(0) = P(1) = \frac{1}{2}$.

Then $K_{00} = K_{11} = \{0\}$, $K_{01} = K_{10} = \{1\}$. Theorem 2 tells us that F is perfectly secure.

The Shift Cipher

We provide $M_0 = K = C_0$ with a group structure, and let $F\colon M_0 \times K \longrightarrow C_0$ be the group composition, hence $f_k(a) = a * k$. The sets

$$K_{ac} = \{k \in K \mid a * k = c\} = \{a^{-1} * c\}$$

each consist of one element only. We let $P(k) = \frac{1}{\#K}$ for all keys $k \in K$. Then F is perfectly secure.

The Examples 0 and 1 are the special cases of the one- or two-element group. Also Examples 2 and 3 will be special cases.

Example 2: The Caesar Cipher. This is the shift cipher on the cyclic group $\Sigma = \mathbb{Z}/n\mathbb{Z}$ of order n.

Hence the Caesar cipher is perfecly secure, *if we encrypt messages of length 1 only and randomly choose an independent new key for each message.*

Example 3: The One-Time Pad. This is the collection of the shift ciphers on the groups $\Sigma^r = M_0$ where $\Sigma = \mathbb{Z}/n\mathbb{Z}$. Messages are texts of length r, and keys are *independently and randomly chosen* letter sequences of the same length r.

Because one has to choose a new key for each message this cipher has its name **One-Time Pad**. Imagine a tear-off calendar where each sheet contains a random letter. After use it is torn off and destroyed.

The One-Time Pad is the prototype of a perfect cipher.

The special case $\Sigma = \{0, 1\}$ gives the binary VERNAM/MAUBORGNE cipher, that is the bitstram encryption with a completely random sequence of key bits.

Counterexample: The Monoalphabetic Substitution. Set $M_0 = \Sigma^r$ and $K = \mathcal{S}(\Sigma)$. For $r = 5$ we saw already that

$$P(\texttt{fruit}|\texttt{XTJJA}) = 0 < q = P(\texttt{fruit}).$$

Therefore the monoalphabetic substitution is not perfect (for $r \geq 2$ and $n \geq 2$). For $r = 1$ it is perfect by Theorem 2 (with $s = (n-1)!$).

4 Density and Redundancy of a Language

SHANNON's theory provides an idea of an unbreakable cipher via the concept of perfection. Moreover it develops the concept of "unity distance" as a measure of the difference to perfection. This concept takes up the observation that the longer a ciphertext, the easier is its unique decryption.

We don't want to develop this theory in a mathematically precise way, but only give a rough impression. For a mathematicly more ambitious approach see [11].

Unique Solution of the Shift Cipher

Let the ciphertext `FDHVDU` be the beginning of a message that was encrypted using a CAESAR cipher. We solved it by exhaustion applying all possible 26 keys in order:

Key	Plaintext	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0	`fdhvdu`	+					
1	`ecguct`	+	+				
2	`dbftbs`	+					
3	`caesar`	+	+	+	+	+	+
4	`bzdrzq`	+					
5	`aycqyp`	+	+				
6	`zxbpxo`	+					
7	`ywaown`	?					
8	`xvznvm`	?					
9	`wuymul`	+	+				
10	`vtxltk`	+					
11	`uswksj`	+	+	?			
12	`trvjri`	+	+				
13	`squiqh`	+	+	+	+		
14	`rpthpg`	+					
15	`qosgof`	+					
16	`pnrfne`	+	+				
17	`omqemd`	+	+				
18	`nlpdlc`	+					
19	`mkockb`	+					
20	`ljnbja`	+					
21	`kimaiz`	+	+	+	?	?	
22	`jhlzhy`	+					
23	`igkygx`	+	+				
24	`hfjxfw`	+					
25	`geiwev`	+	+	+	?		

The flags in this table stand for:

- +: The assumed plaintext makes sense including the t-th letter.
- ?: The assumed plaintext could make sense including the t-th letter but with low probability.

Given the first five letters only one of the texts seems to make sense. We would call this value 5 the "unicity distance" of the cipher.

Mathematical Model

Let us start again with an n-letter alphabet Σ. The "information content" of a letter is $\log_2 n$, for we need $\lceil \log_2 n \rceil$ bits for a binary encoding of all of Σ.

Example For $n = 26$ we have $\log_2 n \approx 4.7$. Thus we need 5 bits for encoding all letters differently. One such encoding is the teleprinter code.

Now let $M \subseteq \Sigma^*$ be a language. Then $M_r = M \cap \Sigma^r$ is the set of "meaningful" texts of length r, and $\Sigma^r - M_r$ is the set of "meaningless" texts. Denote the number of the former by

$$t_r := \# M_r.$$

Then $\log_2 t_r$ is the "information content" of a text of length r or the **entropy** of M_r. This is the number of bits we need for distinguishing the elements of M_r in a binary encoding.

Remark More generally the entropy is defined for a model that assigns the elements of M_r different probabilities. Here we implicitly content ourselves with using a uniform probability distribution.

We could consider the relative frequency of meaningful texts, t_r / n^r, but instead we focus on the **relative information content**,

$$\frac{\log_2 t_r}{r \cdot \log_2 n} :$$

For an encoding of Σ^r we need $r \cdot \log_2 n$ bits, for an encoding of M_r only $\log_2 t_r$ bits. The relative information content is the factor by which we can "compress" the encoding of M_r compared with that of Σ^r. The complimentary portion

$$1 - \frac{\log_2 t_r}{r \cdot \log_2 n}$$

is "redundant".

Usually one relates these quantities to $\log_2 n$, the information content of a single letter, and defines:

Definition 2 (i) The quotient

$$\rho_r(M) := \frac{\log_2 t_r}{r}$$

is called the r**-th density**, the difference $\delta_r(M) := \log_2 n - \rho_r(M)$ is called the r**-th redundancy** of the language M.

(ii) If $\rho(M) := \lim_{r\to\infty} \rho_r(M)$ exists, it is called the **density** of M, and $\delta(M) := \log_2 n - \rho(M)$ is called the **redundancy** of M.

Remarks

1. Since $0 \le t_r \le n^r$, we have $\overline{\lim}\, \rho_r(M) \le \log_2 n$.
2. If $M_r \neq \emptyset$, then $t_r \ge 1$, hence $\rho_r(M) \ge 0$. If $M_r \neq \emptyset$ for almost all r, then $\underline{\lim}\, \rho_r(M) \ge 0$.
3. If $\rho(M)$ exists, then $t_r \approx 2^{r\rho(M)}$ for large r.

For natural languages one knows from empirical observations that $\rho_r(M)$ is (more or less) monotonically decreasing. Therefore density and redundancy exist. Furthermore $t_r \ge 2^{r\rho(M)}$. Here are some empirical values (for $n = 26$):

M	$\rho(M) \approx$	$\delta(M) \approx$
English	1.5	3.2
German	1.4	3.3

The redundancy of English is $\frac{3.2}{4.7} \approx 68\%$ (but [2] says 78%; also see [10]). One expects that an English text (written in the 26 letter alphabet) can be compressed by this factor. The redundancy of German is about $\frac{3.3}{4.7} \approx 70\%$ [10].

5 Unicity Distance

We now apply our findings on the redundancy to the exhaustion of the key space. We don't deal with the expenses but only consider the feasibility. We follow the simplified approach of Hellman.

Assumptions

1. All meaningful texts of length r have the same probability. [Otherwise we get more complicated formulas. For natural languages this assumption is clearly false when r is small. However for large r we might hope that it follows from the usual stochastic conditions.]
2. The densitiy $\rho(M)$ of the language M exists. [Otherwise we could derive only a bound.]
3. All keys $k \in K$ have the same probability and they are $h = \#K$ in number.
4. All encryption functions f_k for $k \in K$ respect the lengths of the texts, or in other words $f(M_r) \subseteq \Sigma^r$.

Now let $c \in \Sigma^r$ be a ciphertext. In general—if all encryption functions f_k are different—it fits h possible plaintexts of length r in Σ^r. By far not all of them are meaningful but only

$$h \cdot \frac{t_r}{n^r} \approx \frac{h \cdot 2^{r\rho(M)}}{2^{r \cdot \log_2 n}} = h \cdot 2^{-r\delta(M)}.$$

We expect a unique solution in M_r if

$$h \cdot 2^{-r\delta(M)} \leq 1, \quad \log_2 h - r\delta(M) \leq 0, \quad r \geq \frac{\log_2 h}{\delta(M)},$$

at least if all encryption functions f_k are different; otherwise we should replace $\log_2 h$ with $d = d(F)$, the effective key length of the cipher F.

This motivates the following definition:

Definition 3. For a cipher F with effective key length $d(F)$ defined on a language M of redundancy $\delta(M)$ we call

$$\mathrm{UD}(F) := \frac{d(F)}{\delta(M)}$$

the **unicity distance**.

Examples

We always assume the alphabet $\Sigma = \{\mathtt{A}, \ldots, \mathtt{Z}\}$ with $n = 26$, and the language $M =$ "English".

1. For the shift cipher we have $d = \log_2 26$, UD $\approx 4.7/3.2 \approx 1.5$, not about 5 as suspected in the introductory example. This deviation might be due to the many inexact steps in the derivation. In particular for small r the approximation $t_r \approx 2^{r\rho(M)}$ is very inexact.

2. For the monoalphabetic substitution we have $d \approx 88.4$, UD $\approx 88.4/3.2 \approx 27.6$. This result is in good concordance with empirical observations on the solvability of monoalphabetic cryptograms.

3. For the Trithemius-Bellaso cipher with period l we have $d \approx 4.7 \cdot l$, UD $\approx 1.5 \cdot l$.

4. For Porta's disk cipher we have $d \approx 88.4 + 4.7 \cdot l$, UD $\approx 27.6 + 1.5 \cdot l$.

5. For the general polyalphabetic substitution with period l and independent alphabets $d \approx 122 \cdot l$, UD $\approx 38 \cdot l$.

6. For the One-Time Pad over the group $G = \Sigma$ we have $M = K = C = \Sigma^*$, hence $\#K = \infty$. However it makes sense to interpret $d_r/\delta_r = r \cdot \log_2 n/0 = \infty$ as unicity distance.

6 Cryptological Applications

The unicity distance is a very coarse measure of the quality of a cipher. In modern cryptology it is almost never used. For an attack with known plaintext it is meaningless (except for perfect ciphers where it is ∞).

A large unicity distance is achieved by:

- a large key space,
- lowering the redundancy of the plaintext language, for example by compression.

Application 1: PORTA's disk cipher is not so much stronger than the TRITHEMIUS-BELLASO cipher because the unicity distance is greater only by the constant summand 27.6. For a longer period the complication by permuting the primary alphabet effects not much additional security.

Application 2: Another application of SHANNON's theory is to running text encryption. The cryptanalysis must derive two meaningful plaintexts of total length $2r$ from a ciphertext of length r. This can work only for a language of redundancy at least 50%.

More generally consider a q-fold running text encryption with q independent keytexts. If cryptanalysis is feasible, then meaningful plaintext of total length $(q+1) \cdot r$ is excavated from a ciphertext of length r. We conclude that the redundancy of the language is at least $\geq \frac{q}{q+1}$. Because the redundancy of German, 70%, is smaller than $\frac{3}{4}$ we conclude that a triple running text encryption is secure. For English that has somewhat less redundancy even a double running text encryption seems to be secure.

Application 3: The unicity distance may serve as an indication for how much ciphertext corresponding to a single key may be known to the enemy without being of use. Or in other words: How often the key must change.

A general short summary of SHANNON's theory consists of the rule: *A necessary condition for the solvability of a cipher is that "information content of the ciphertext + redundancy of the plaintext language" ≥ "information content of the plaintext + information content of the key".*

References

[1] F. L. Bauer, *Decrypted Secrets; Methods and Maxims of Cryptology.* Springer, Berlin 1997.

[2] C. A. Deavours, Unicity points in cryptanalysis. Cryptologia 1 (1977), 469–684.

[3] C. A. Deavours, L. Kruh, *Machine Cryptography and Modern Cryptanalysis.* Artech House, Norwood 1985.

[4] W. F. Friedman, *The Riverbank Publications Volume 1* (contains Publications No. 15, 16, 17, and 18). Aegean Park Press, Laguna Hills 1979.

[5] R. Ganesan, A. T. Sherman, *Statistical Techniques for Language Recognition: An Introduction and Guide for Cryptanalysts.* Cryptologia 17 (1993), 321–366.

[6] R. Ganesan, A. T. Sherman, *Statistical Techniques for Language Recognition: An Empirical Study Using Real and Simulated English.* Cryptologia 18 (1994), 289–331.

[7] A. M. Gleason, *Elementary Course in Probability for the Cryptanalyst.* Aegean Park Press, Laguna Hills 1985.

[8] M. E. Hellman, An extension of the Shannon theory approach to cryptography. IEEE Trans Information Theory 23 (1977), 289–294.

[9] A. M. Jaglom, I. M, Jaglom, *Wahrscheinlichkeit und Information.* VEB Deutscher Verlag der Wissenschaften, Berlin 1967.

[10] H. Jürgensen, Language redundancy and the unicity point. Cryptologia 7 (1983), 37–48.

[11] H. Jürgensen, D. E. Matthews, Some results on the information theoretic analysis of cryptosystems. Crypto 83, 303–356.

[12] S. Kullback, *Statistical Methods in Cryptanalysis.* Aegean Park Press, Laguna Hills 1976.

[13] J. Reeds, Entropy calculations and particular methods of cryptanalysis. Cryptologia 1 (1977), 235–254.

[14] C. E. Shannon, A mathematical theory of communication. Bell System Technical Journal 27 (1948), 379–423, 623–656.

[15] C. E. Shannon, Communication theory of secrecy systems. Bell System Technical Journal 28 (1949), 656–715.

[16] C. E. Shannon, The entropy of printed english. Bell System Technical Journal 30 (1941), 50–64.

[17] A. Sinkov, *Elementary Cryptanalysis*. The Mathematical Association of America, Washington, 1966.