A Large–Deviations Notion of Perfect Secrecy

Neri Merhav*

March 2, 2011

Abstract

We consider the Shannon cipher system with a variable key rate, and study the necessary and sufficient conditions for perfect secrecy in the sense that the exponential rate of the probability of breaking into the system would not be improved by observing the cryptogram. For a memoryless plaintext source, we derive achievable lower bounds on the number of key bits needed for *almost every plaintext sequence* in every type class. The corresponding minimum achievable average key rate turns out to be the negative logarithm of the probability of the most likely plaintext letter, which is in general, smaller than the entropy.

Index Terms: Shannon cipher system, cryptography, cryptanalysis.

*Department of Electrical Engineering, Technion -I.I.T., Haifa 32000, Israel. E-mail: [merhav@ee.technion.ac.il].

1 Introduction

In the classical Shannon-theoretic approach to cryptology [5], the security of cipher systems is traditionally measured in terms of the equivocation, that is, the conditional entropy of the plaintext (or the key) given the cryptogram. As is well known (see, e.g., [3]), this conditional entropy can be at most as large as the rate of the purely random key stream. Thus, perfect *theoretical secrecy* is attainable if and only if the key rate is at least as large as the message rate. Other, less pessimistic, information–theoretic notions of security were also proposed. For example, Hellman [2] proposed to measure the degree of security of a cryptosystem in terms of the expected number of *spurious messages*, i.e., the expected number of plaintext-key combinations that may explain the given cryptogram. The assumption in [2] is that the number of meaningful messages of a given length n within the language of the source, is very small compared to the total number of possible n-vectors. Another interesting definition is Maurer's conditionally–perfect security as well as his construction of a low key-rate randomized cipher [4], which is secure (in Shannon's sense) provided that a certain event occurs whose probablity is high unless the cryptanalyzer performs a computationally infeasible task.

In this correspondence, we propose to define secrecy in a large–deviations sense: A cryptosystem will be considered secure if the presence of the cryptogram does not improve the exponential rate of the probability of breaking into the system, namely, deciphering the correct message without access to the key. More specifically, we consider Shannon's model of a secrecy system [5], where a plaintext message $\mathbf{X} = (X_1, \ldots, X_n)$, emitted by a discrete memoryless source P, is to be communicated as securely as possible from a transmitter to a legitimate receiver. The transmitter and receiver have access to a common key string $\mathbf{U}$ of purely random bits, whose length $K = K(\mathbf{X})$ may depend on $\mathbf{X}$. The transmitter generates a cryptogram

$$\mathbf{Y} = \phi(\mathbf{X}, \mathbf{U})$$

and sends it over a public channel to the receiver. The cryptogram $\mathbf{Y}$ is a string (possibly, of variable length) over an alphabet that is not necessarily the same as the source alphabet. The encryption function is invertible given the key in the sense that there exists an inverse decryption function

$$\mathbf{X} = \phi^{-1}(\mathbf{Y}, \mathbf{U})$$

to be used by the legitimate receiver who observes both $\mathbf{Y}$ and $\mathbf{U}$. An enemy wiretapper, who knows the encryption function ϕ (and hence also the decryption function ϕ^{-1}) and the statistics of the plaintext source, but not the key itself, aims at decrypting $\mathbf{X}$ from the observed cryptogram $\mathbf{Y}$ only.

Clearly, the probability of correctly guessing the plaintext, based only on knowing the probability mass function P, but without the cryptogram, is given by

$$\max_{\mathbf{x}} P(\mathbf{x}) = [\max_{x} P(x)]^n = 2^{-n\Gamma_S},$$

where x is a single plaintext symbol, $\mathbf{x}$ is a plaintext string of length n, and

$$\Gamma_S \stackrel{\triangle}{=} -\log \max_{x} P(x).$$

The question we address is then the following: How many key bits, $K(\mathbf{X})$, should be used to encrypt every $\mathbf{X}$ so as to guarantee that the probability P_C of correctly deciphering $\mathbf{X}$ by an eavesdropper who observes $\mathbf{Y}$ (but not $\mathbf{U}$), will continue to decay at the exponential rate of $2^{-n\Gamma_S}$?

Our main result is that a necessary condition for this to be the case is that for *almost* every sequence $\mathbf{X}$ in every type class T_Q (where Q is the empirical probability mass function of single letters associated with $\mathbf{X}$), $K(\mathbf{X})$ must be essentially at least as large as

$$n[\Gamma_S - D(Q\|P)]_+,$$

where $[u]_+ \stackrel{\triangle}{=} \max\{u, 0\}$. On the other hand, it is easy to show that there exists a simple cipher system with $K(\mathbf{X}) \approx n[\Gamma_S - D(Q\|P)]_+$ for *all* $\mathbf{X}$, which satisfies the above–mentioned security requirement, namely, $P_C \sim 2^{-n\Gamma_S}$. Therefore, essentially the same condition is sufficient as well.

An immediate consequence of this result is that the needed key rate $R(\mathbf{X}) = K(\mathbf{X})/n$ for each $\mathbf{X}$ essentially never exceeds Γ_S, which is in turn less than or equal to the entropy of the source H_S (with equality when all letters are equally likely). Therefore, this notion of secrecy is less pessimistic than Shannon's notion of equivocation.

One might argue, on the other hand, that this may be even overly optimistic, because if the eavsdropper deciphers correctly as many as 99% of the plaintext symbols (but not the remaining 1%), this is considered as a failure from the viewpoint of breaking into the system. In some applications, this assumption is indeed well–justified, for example, when

$\mathbf{X}$ is a secret personal verification message, like a password of a computer account or a bank account number accessed via the Internet. In other applications, a more plausible approach would be to adopt a criterion that provides better protection even when the eavesdropper's estimate $\hat{\mathbf{X}}$ is only close to the true message under some fidelity criterion $\rho(\mathbf{X}, \hat{\mathbf{X}})$ (see, e.g., [6]). For example, instead of maintaining the exponential rate of P_C as proposed above, one might be interested to maintain the exponential rate of the probability of the event $\rho(\mathbf{X}, \hat{\mathbf{X}}) \leq nD$ (for a given D) at the same level as in the absence of a cryptogram. We have not pursued this direction in this work.

2 Definitions and Notation Conventions

Throughout this correspondence, scalar random variables will be denoted by capital letters while their sample values will be denoted by the respective lower case letters. A similar convention will apply to random vectors and their sample values, which will be denoted by boldface letters. Thus, for example, if $\mathbf{X}$ denotes a random vector $(X_1, ..., X_n)$, then $\mathbf{x} = (x_1, ..., x_n)$ would designate a specific realization of $\mathbf{X}$.

The plaintext message will be assumed to be drawn from a discrete memoryless source (DMS) with a finite alphabet $\mathcal{X}$ and probability mass function (PMF) $P = \{P(x),\ x \in \mathcal{X}\}$. The probability of a vector $\mathbf{x}$, will be denoted $P(\mathbf{x})$, which is given by $\prod_{i=1}^n P(x_i)$. The nth order Cartesian power of $\mathcal{X}$, that is, the space of all n-vectors over $\mathcal{X}$, will be denoted by $\mathcal{X}^n$. The probability of an event $A \subseteq \mathcal{X}^n$ will be denoted by $P(A)$ or $\Pr\{A\}$. We shall use the letter Q to denote a generic DMS over the alphabet $\mathcal{X}$, and use the same notational conventions as for P.

For a DMS Q, we recall that the Shannon entropy is given by

$$H(Q) = -\sum_{x \in \mathcal{X}} Q(x) \log Q(x), \tag{1}$$

where logarithms throughout the sequel are taken to the base 2. The relative entropy between Q and P is defined as

$$D(Q\|P) = \sum_{x \in \mathcal{X}} Q(x) \log \frac{Q(x)}{P(x)}. \tag{2}$$

For a given source vector $\mathbf{x} \in \mathcal{X}^n$, the empirical probability mass function (EPMF) is the vector $Q_{\mathbf{x}} = \{Q_{\mathbf{x}}(a), a \in \mathcal{X}\}$, where $Q_{\mathbf{x}}(a) = n_{\mathbf{x}}(a)/n$, $n_{\mathbf{x}}(a)$ being the number of occurrences of the letter a in the vector $\mathbf{x}$. The set of all EPMF's of vectors in $\mathcal{X}^n$, that is, rational PMF's with denominator n, will be denoted by $\mathcal{Q}_n$. The type class $T_{\mathbf{x}}$ of a vector $\mathbf{x}$ is the set of all vectors $\mathbf{x}' \in \mathcal{X}^n$ such that $Q_{\mathbf{x}'} = Q_{\mathbf{x}}$. When we need to attribute a type class to a certain rational PMF $Q \in \mathcal{Q}_n$ rather than to a sequence in $\mathcal{X}^n$, we shall use the notation T_Q. It is well-known [1] that the number of type classes of n-vectors is bounded by $(n+1)^{|\mathcal{X}|-1}$, where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$. The standard reference about the method of types is the book by Csiszár and Körner [1]. Finally, throughout the sequel, $O(n)$ desginates a quantity that grows asymptotically linearly with n, i.e., $O(n)/n$ tends to a constant.

3 Main Results

For a given cipher system ϕ, let $P(\mathbf{y}|\mathbf{x})$ denote the induced conditional probability of the cryptogram $\mathbf{y}$ given the plaintext $\mathbf{x}$. Similarly, let $P(\mathbf{x},\mathbf{y}) = P(\mathbf{x})P(\mathbf{y}|\mathbf{x})$ denote the joint probability mass function, and let $P(\mathbf{x}|\mathbf{y})$ and $P(\mathbf{y})$ be the induced conditional probability of $\mathbf{x}$ given $\mathbf{y}$ and the marginal of $\mathbf{y}$, respectively.

Since the best estimator of $\mathbf{x}$ given $\mathbf{y}$ (in the sense of maximizing P_C) is given by

$$\hat{\mathbf{x}} = \text{argmax}_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}),$$

then the probability of optimum correct decryption of $\mathbf{X}$ in the presence of the cryptogram is clearly given by

$$P_C = \sum_{\mathbf{y}} P(\mathbf{y}) \max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}) = \sum_{\mathbf{y}} \max_{\mathbf{x}} P(\mathbf{x},\mathbf{y}). \tag{3}$$

Our first result tells that for P_C to decay as fast as 2^{-nE}, almost all sequences within every type class, T_Q, must be encrypted using essentially at least $n[E - D(Q\|P)]_+$ random bits. Perfect security then corresponds to the special case where $E = \Gamma_S$.

Theorem 1 *For a given $E > 0$, if $P_C \le 2^{-nE}$, then for every type class T_Q, the following holds: For every $\epsilon > 0$,*

$$|T_Q \cap \{\mathbf{x}: \ K(\mathbf{x}) \le n([E - D(Q\|P)]_+ - \epsilon)\}| \le 2^{O(\log n) - n\epsilon}|T_Q|. \tag{4}$$

Proof. First observe that for type classes $\{T_Q\}$ where $D(Q\|P) \ge E$, the assertion of the theorem is trivial since the set $\{\mathbf{x}: \ K(\mathbf{x}) < 0\}$ is empty. Consider then an arbitrary type class for which $D(Q\|P) < E$, in which case, the operation $[\cdot]_+$ is neutral. By the same token, if $\epsilon > E - D(Q\|P)$, the assertion of the theorem is again trivial. Assume then that $0 < \epsilon \leq E - D(Q\|P)$.

For a given cipher system ϕ, let $\phi^{-1}(\mathbf{y})$ denote the set of all $\mathbf{x}$ for which there exists a key string $\mathbf{u} = (u_1, \ldots, u_{K(\mathbf{x})})$ such that $\mathbf{y} = \phi(\mathbf{x}, \mathbf{u})$. Also, for a non–negative integer $s \in \{0, 1, \ldots, \lceil n \log |\mathcal{X}| \rceil\}$, let

$$T_Q^s = T_Q \cap \{\mathbf{x}: \; K(\mathbf{x}) = s\}.$$

Note that the number of distinct sets $\{T_Q^s\}$ is upper bounded by

$$M_n \stackrel{\triangle}{=} |\mathcal{Q}_n| \cdot (\lceil n \log |\mathcal{X}| \rceil + 1) \leq (n \log |\mathcal{X}| + 2) \cdot (n+1)^{|\mathcal{X}|-1},$$

which is a polynomial in n. Now,

$$\begin{aligned}
P_C &= \sum_{\mathbf{y}} P(\mathbf{y}) \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y})} P(\mathbf{x}|\mathbf{y}) \\
&= \sum_{\mathbf{y}} P(\mathbf{y}) \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y})} P(\mathbf{x}, T_Q^s|\mathbf{y}) \\
&= \sum_{\mathbf{y}} P(\mathbf{y}) \max_{T_Q^s} \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y})} P(\mathbf{x}, T_Q^s|\mathbf{y}) \\
&= \sum_{\mathbf{y}} P(\mathbf{y}) \max_{T_Q^s} \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y}) \cap T_Q^s} P(T_Q^s|\mathbf{y}) P(\mathbf{x}|T_Q^s, \mathbf{y}) \\
&= \sum_{\mathbf{y}} P(\mathbf{y}) \max_{T_Q^s} P(T_Q^s|\mathbf{y}) \cdot \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y}) \cap T_Q^s} P(\mathbf{x}|T_Q^s, \mathbf{y}) \\
&\geq \frac{1}{M_n} \sum_{\mathbf{y}} P(\mathbf{y}) \sum_{T_Q^s:\; P(T_Q^s|\mathbf{y})>0} P(T_Q^s|\mathbf{y}) \cdot \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y}) \cap T_Q^s} P(\mathbf{x}|T_Q^s, \mathbf{y}) \\
&= \frac{1}{M_n} \sum_{\mathbf{y}} \sum_{T_Q^s:\; P(T_Q^s|\mathbf{y})>0} \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y}) \cap T_Q^s} P(T_Q^s) P(\mathbf{x}|T_Q^s) P(\mathbf{y}|\mathbf{x}) \\
&= \frac{1}{M_n} \sum_{\mathbf{y}} \sum_{T_Q^s:\; P(T_Q^s|\mathbf{y})>0} \max_{\mathbf{x} \in \phi^{-1}(\mathbf{y}) \cap T_Q^s} P(T_Q^s) \cdot \frac{1}{|T_Q^s|} \cdot 2^{-s} \\
&= \frac{1}{M_n} \sum_{\mathbf{y}} \sum_{T_Q^s:\; P(T_Q^s|\mathbf{y})>0} \frac{P(T_Q^s)}{|T_Q^s|} \cdot 2^{-s} \\
&= \frac{1}{M_n} \sum_{T_Q^s} \sum_{\mathbf{y}:\; P(T_Q^s|\mathbf{y})>0} \frac{P(T_Q^s)}{|T_Q^s|} \cdot 2^{-s} \\
&= \frac{1}{M_n} \sum_{T_Q^s} |\{\mathbf{y}: \; P(T_Q^s|\mathbf{y}) > 0\}| \cdot \frac{P(T_Q^s)}{|T_Q^s|} \cdot 2^{-s} \\
&\geq \frac{1}{M_n} \sum_{T_Q^s} |T_Q^s| \cdot \frac{P(T_Q^s)}{|T_Q^s|} \cdot 2^{-s} \\
&\geq \frac{1}{M_n} \max_{T_Q^s} P(T_Q^s) \cdot 2^{-s}, \qquad (5)
\end{aligned}$$

where in the second to the last inequality we have used the fact that the set $\{\mathbf{y}:\ P(T_Q^s|\mathbf{y})>0\}$ is actually identical to the set $\{\mathbf{y}=\phi(\mathbf{x},\mathbf{u}):\ \mathbf{u}\in\{0,1\}^{K(\mathbf{x})},\ \mathbf{x}\in T_Q^s\}$, whose cardinality cannot be smaller than $|T_Q^s|$ since ϕ must be invertible given $\mathbf{u}$. It now follows from the hypothesis of the theorem that for every Q and s

$$2^{-nE}\geq\frac{1}{M_n}P(T_Q^s)\cdot 2^{-s}\geq\frac{1}{M_n^2}\cdot\frac{|T_Q^s|}{|T_Q|}\cdot 2^{-nD(Q\|P)-s} \tag{6}$$

or, equivalently,

$$|T_Q^s|\leq 2^s M_n^2|T_Q|\cdot 2^{-n[E-D(Q\|P)]}. \tag{7}$$

Thus, for every non-negative z,

$$\begin{aligned}|T_Q\cap\{\mathbf{x}:\ K(\mathbf{x})\leq z\}| &= \sum_{s=0}^{\lfloor z\rfloor}|T_Q^s|\\ &\leq M_n^2\cdot|T_Q|\cdot 2^{-n[E-D(Q\|P)]}\cdot\sum_{s=0}^{\lfloor z\rfloor}2^s\\ &\leq 2M_n^2\cdot|T_Q|\cdot 2^{-n[E-D(Q\|P)]}\cdot 2^z,\end{aligned} \tag{8}$$

and the proof is completed by setting $z=n[E-D(Q\|P)-\epsilon]$ and using the fact that $\log M_n=O(\log n)$. □

We next demonstrate a conceptually simple cipher system for which

$$K(\mathbf{x})\leq n[\Gamma_S-D(Q_{\mathbf{x}}\|P)]_+$$

for all $\mathbf{x}$, while keeping P_C no larger than the exponential order of $2^{-n\Gamma_S}$.

This cipher system works as follows: First, compress $\mathbf{x}$ losslessly into a binary vector of two fields. The first field encodes the index of the type class $T_{\mathbf{x}}$ using $O(\log n)$ bits, and the second field contains the index of $\mathbf{x}$ within $T_{\mathbf{x}}$ using $\lceil\log|T_{\mathbf{x}}|\rceil$ bits. If $D(Q_{\mathbf{x}}\|P)<\Gamma_S$, encrypt the first[1] $K(\mathbf{x})=n[\Gamma_S-D(Q_{\mathbf{x}}\|P)]$ bits of the second field by applying a bit–by–bit XOR operation with the same number of key bits. The cryptogram $\mathbf{y}$ is then the partially encrypted binary codevector for $\mathbf{x}$. If $D(Q_{\mathbf{x}}\|P)\geq\Gamma_S$, do not encrypt at all and let $\mathbf{y}$ be just the compressed bit string of $\mathbf{x}$.[2]

To see why this scheme gives the desired behavior of P_C, first observe that the contribution of type classes for which $D(Q\|P)>\Gamma_S$ can be neglected because their probabilities

[1] It is easy to see that $n[\Gamma_S-D(Q_{\mathbf{x}}\|P)]$ never exceeds $nH(Q_{\mathbf{x}})$, which is the approximate size of the second field.

[2] While this scheme formally achieves the goal of attaining Γ_S, the fact that some plaintexts are not protected at all, may be objectionable. A simple modification could be to encrypt at least a small fraction of the compressed bits of every such plaintext anyhow.

decay faster than the target exponential rate of $2^{-n\Gamma_S}$. Confining then attention to the remaining type classes, repeat the chain of equations (5) with the above described scheme in mind, where in this case, $\{T_Q^s\}$ are all empty except for $s = s(Q) \stackrel{\triangle}{=} n[\Gamma_S - D(Q\|P)]$, as $T_Q^{s(Q)}$ is populated by the entire type class T_Q. Now, the first and the last inequalities in (5) are always exponentially tight. Thus the only possible cause of lack of exponential tightness in eqs. (5) might be the second to the last inequality, which is nevertheless tight as well (according to the explanation that follows (5)) if our scheme satisfies

$$|\{\mathbf{y} = \phi(\mathbf{x}, \mathbf{u}): \ \mathbf{u} \in \{0,1\}^{K(\mathbf{x})}, \ \mathbf{x} \in T_Q\}| = |T_Q|.$$

But this is clearly the case, because the left–hand side corresponds to all $|T_Q|$ possible binary vectors in the second field.

Thus, according to (6), P_C is of the exponential order of

$$\max_Q 2^{-nD(Q\|P)-s_Q} = \max_Q 2^{-nD(Q\|P)-n[\Gamma_S - D(Q\|P)]} = 2^{-n\Gamma_S}. \tag{9}$$

4 Discussion

The last few lines of the proof of Theorem 1 suggest that, in fact, a somewhat more general and more refined argument can be made: If P_C decays at the exponential order of 2^{-nE}, then for every type class T_Q, the fraction of sequences that may be encrypted by no more than nR random bits (assuming $0 < R < E - D(Q\|P)$) essentially cannot exceed $2^{-n[E-D(Q\|P)-R]}$. This actually characterizes a bound on the best achievable *distribution* of key length assignments within each type class.

Another interesting variant of our problem corresponds to the case where the plaintext source P is unknown to the encrypter (except for the fact of being memoryless), but we would like to guarantee that P_C continues to decay at the exponential rate of $2^{-n\Gamma_S}$ for every memoryless P, and even if the cryptanalyzer knows the statistics. It is easy to show that the derivations above extend straightforwardly and the minimum number of key bits needed (for most) sequences within each type class T_Q is given by

$$\max_P [\Gamma_S(P) - D(Q\|P)].$$

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[2] M. E. Hellman, “An extension of the Shannon theory approach to cryptography,” *IEEE Trans. Inform. Theory,* vol. IT-23, no. 3, pp. 289-294, May 1977.

[3] J. L. Massey, “An introduction to contemporary cryptology,” *Proc. IEEE*, vol. 26, no. 5, pp. 533-549, May 1988.

[4] U. M. Maurer, “Conditionally-perfect-secrecy and a provably-secure randomized cipher,” *J. Cryptology*, vol. 5, no. 53–66, 1992.

[5] C. E. Shannon, “Communication theory of secrecy systems,” *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 565-715, Oct. 1949.

[6] H. Yamamoto, “Rate-distortion theory for the Shannon cipher system,” *IEEE Trans. Inform. Theory,* vol. IT-43, pp. 827-835, May 1997.