[redacted] 17

L-1781-B (4-44)

COVER SHEET FOR TECHNICAL MEMORANDA

RESEARCH DEPARTMENT

SUBJECT: A Mathematical Theory of Cryptography - Case 20878 (4)

ROUTING:
1 – H.B-HF-Case Files
2 – CASE FILES
3 – J. W. McRae
4 – L. Espenschied
5 – H. S. Black
6 – F. B. Llewellyn
7 – H. Nyquist
8 – B. M. Oliver
9 – R. K. Potter
10 – C. B. H. Feldman
11 - R. C. Mathes
12 - R. V. L. Hartley
13 - J. R. Pierce
14 - H. W. Bode
15 - R. L. Dietzold
16 - L. A. MacCall
17 - W. A. Shewhart
18 - S. A. Schelkunoff
19 - C. E. Shannon
20 - Dept. 1000 Files

MM- 45-110-92
DATE September 1, 1945
AUTHOR C. E. Shannon
INDEX NO. P 0.4

[redacted]

~~ABSTRACT~~

DOWNGRADED AT 3 YEAR INTERVALS
DECLASSIFIED AFTER 12 YEARS
DOD DIR 5200.10

ABSTRACT

A mathematical theory of secrecy systems is developed. Three main problems are considered. (1) A logical formulation of the problem and a study of the mathematical structure of secrecy systems. (2) The problem of "theoretical secrecy," i.e., can a system be solved given unlimited time and how much material must be intercepted to obtain a unique solution to cryptograms. A secrecy measure called the "equivocation" is defined and its properties developed. (3) The problem of "practical secrecy." How can systems be made difficult to solve, even though a solution is theoretically possible.

THIS DOCUMENT CONTAINS INFORMATION AFFECTING THE NATIONAL DEFENSE OF THE UNITED STATES WITHIN THE MEANING OF THE ESPIONAGE LAWS, TITLE 18 U.S.C. SECTIONS 793 AND 794. ITS TRANSMISSION OR THE REVELATION OF ITS CONTENTS IN ANY MANNER TO AN UNAUTHORIZED PERSON IS PROHIBITED BY LAW.

BEST COPY AVAILABLE

[redacted]

[redacted]

A Mathematical Theory of Cryptography - Case 20878 (4)

MM-45-110-92

September 1, 1945

Index P0.4

MEMORANDUM FOR FILE

DOWNGRADED AT 3 YEAR INTERVALS
DECLASSIFIED AFTER 12 YEARS
DOD DIR 5200.10

Introduction and Summary

In the present paper a mathematical theory of cryptography and secrecy systems is developed. The entire approach is on a theoretical level and is intended to complement the treatment found in standard works on cryptography.* There, a detailed study is made of the many standard types of codes and ciphers, and of the ways of breaking them. We will be more concerned with the general mathematical structure and properties of secrecy systems.

The presentation is mathematical in character. We first define the pertinent terms abstractly and then develop our results as lemmas and theorems. Proofs which do not contribute to an understanding of the theorems have been placed in the appendix.

The mathematics required is drawn chiefly from probability theory and from abstract algebra. The reader is assumed to have some familiarity with these two fields. A knowledge of the elements of cryptography will also be helpful although not required.

The treatment is limited in certain ways. First, there are two general types of secrecy system; (1) concealment systems, including such methods as invisible ink, concealing a message in an innocent text, or in a fake covering cryptogram, or other methods in which the existence of the message is concealed from the enemy; (2) "true" secrecy systems where the meaning of the message is concealed by cipher, code, etc., although its existence is not hidden. We consider only the second type--concealment systems are more of a psychological than a mathematical problem. Secondly, the treatment is limited to the case of discrete information, where the information to be enciphered consists of a sequence of discrete symbols, each chosen from a finite set. These symbols may be letters in a

---

*See, for example, H.F.Gaines, "Elementary Cryptanalysis", or M. Givierge, "Cours de Cryptographie."

THIS DOCUMENT CONTAINS INFORMATION AFFECTING THE NATIONAL DEFENSE OF THE UNITED STATES WITHIN THE MEANING OF THE ESPIONAGE LAWS, TITLE 18 U.S.C., SECTIONS 793 AND 794. ITS TRANSMISSION OR THE REVELATION OF ITS CONTENTS IN ANY MANNER TO AN UNAUTHORIZED PERSON IS PROHIBITED BY LAW.

language, words of a language, amplitude levels of a "quantized" speech or video signal, etc., but the main emphasis and thinking has been concerned with the case of letters. A preliminary survey indicates that the methods and analysis can be generalized to study continuous cases, and to take into account the special characteristics of speech secrecy systems.

The paper is divided into three parts. The main results of these sections will now be briefly summarized. The first part deals with the basic mathematical structure of language and of secrecy systems. A language is considered for cryptographic purposes to be a stochastic process which produces a discrete sequence of symbols in accordance with some system of probabilities. Associated with a language there is a certain parameter D which we call the redundancy of the language. D measures, in a sense, how much a text in the language can be reduced in length without losing any information. As a simple example, if each word in a text is repeated a reduction of 50 per cent is immediately possible. Further reductions may be possible due to the statistical structure of the language, the high frequencies of certain letters or words, etc. The redundancy is of considerable importance in the study of secrecy systems.

A secrecy system is defined abstractly as a set of transformations of one space (the set of possible messages) into a second space (the set of possible cryptograms). Each transformation of the set corresponds to enciphering with a particular key and the transformations are supposed reversible (non-singular) so that unique deciphering is possible when the key is known.

Each key and therefore each transformation is assumed to have an a priori probability associated with it--the probability of choosing that key. The set of messages or message space is also assumed to have a priori probabilities for the various messages, i.e., to be a probability or measure space.

In the usual cases the "messages" consist of sequences of "letters." In this case as noted above the message space is represented by a stochastic process which generates sequences of letters according to some probability structure.

These probabilities for various keys and messages are actually the enemy cryptanalyst's a priori probabilities for the choices in question, and represent his a priori knowledge of the situation. To use the system a key is first selected and sent to the receiving point. The choice of a key determines a particular transformation in the set forming the system. Then a message is selected and the particular transformation applied to this message to produce a cryptogram. This cryptogram is

transmitted to the receiving point by a channel that may be intercepted by the enemy. At the receiving end the inverse of the particular transformation is applied to the cryptogram to recover the original message.

If the enemy intercepts the cryptogram he can calculate from it the a posteriori probabilities of the various possible messages and keys which might have produced this cryptogram. This set of a posteriori probabilities constitutes his knowledge of the key and message after the interception.* The calculation of these a posteriori probabilities is the generalized problem of cryptanalysis.

As an example of these notions, in a simple substitution cipher with random key there are 26! transformations, corresponding to the 26! ways we can substitute for 26 different letters. These are all equally likely and each therefore has an a priori probability 1/26!. If this is applied to "normal English" the cryptanalyst being assumed to have no knowledge of the message source other than that it is English the a priori probabilities of various messages of N letters are merely their frequency in normal English text.

If the enemy intercepts N letters of cryptogram in this system his probabilities change. If N is large enough (say 50 letters) there is usually a single message of a poste probability nearly unity, while all others have a total proba bility nearly zero. Thus there is an essentially unique "sol tion" to the cryptogram. For N smaller (say N = 15) there wi usually be many messages and keys of comparable probability, with no single one nearly unity. In this case there are mult "solutions" to the cryptogram.

Considering a secrecy system to be a set of transformations of one space into another with definite probabilit associated with each transformation, there are two natural co bining operations which produce a third system from two given systems. The first combining operation is called the product operation and corresponds to enciphering the message with the first system R and enciphering the resulting cryptogram with system S, the keys for R and S being chosen independently. This total operation is a secrecy system whose transformation consist of all the products (in the usual sense of products o transformations) of transformations in S with transformations in R. The probabilities are the products of the probabilitie for the two transformations.

The second combining operation is "weighted additio

$$T = pR + qS \qquad p + q = 1$$

---

*"Knowledge" is thus identified with a set of propositions ha associated probabilities. We are here at variance with the doctrine often assumed in philosophical studies which consid knowledge to be a set of propositions which are either true false.