Preface to Claude Shannon's
A Mathematical Theory of Cryptography

Whitfield Diffie

December 2015

In 1948, Claude Shannon published the paper "A Mathematical Theory of Communication" which is justly seen as the foundation of modern information theory. A year later, he published a second paper, "Communication Theory of Secrecy Systems" applying information theory to the study of cryptography. The order of presentation and the broad applicability of information theory to communication leave the impression that Shannon developed information theory — presumably to support the work of his employer Bell Telephone Laboratories — first and then applied the new theory to cryptography. In fact, the situation is quite the reverse.

In 1945, Shannon wrote a technical report entitled "A Mathematical Theory of Cryptography" that is at the root of both the 1948 and 1949 papers. This report, which was classified CONFIDENTIAL at the time it was written, has the feel of a what-can-I-contribute-to-the-war-effort undertaking of which there must have been many. If that was its intent, it may have been the most important paper to have arisen from that motivation. It contains all of the content of the 1949 paper and much of the content of the 1948 paper. The early pages of this work make clear that Shannon is developing the information theory concepts in support of cryptography and in a footnote on page eight he says "It is intended to develop these results in coherent fashion in a forthcoming memorandum on the transmission of information."

Some of the examples have proven quite "durable" and were carried forward not only into the formal publications but in the work of Shannon's colleagues. Shannon gives a number of examples of pseudo-English words and phrases generated by Markov models. In Section 3, we read

> To construct [the second order approximation] for example one opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter recorded, etc. A similar process was used for [approximations up to order three] ... It would be interesting if further approximations could be constructed, but the labor involved becomes enormous at the next stage.

In fact, the task is not only computationally infeasible with the computing resources available at the time, it can be seen with those available today that it does not work with even a very large book. Going much further would requre a far larger sample. Such sample's are available today but the effort of processing

them would be better expended in building the probablistic transition tables rather than constructing examples.

The difficulty of creating even the examples of Markov-process English presumably explains why the same examples are repeated verbatim in John Pierce's book *Symbols, Signals and Noise* published a few years later.

In preparing a clean and readable version of "A Mathematical Theory of Cryptography," we have had both the luxury and the annoyance of having two typescripts from which to work. One came from the National Security Agency R5 library and was lent to me decades ago. The other was part of a collection called "Claude Elwood Shannon: Miscellaneous Writings" edited by Neil Sloane and Aaron Wyner, which consists primarily or entirely of items that they omitted from their more formal collection *Claude E. Shannon: Collected Papers*. The failing of the NSA typescript is being too light, while that of the Archive typescript is being both too dark and often a bit off the page on the right. The two versions carry the same dates and version numbers, so we have been unable to determine which is the earlier and which the later. Unfortunately, the versions differ from each other in poth pagination and line breaking. We have chosen the pagination of the Archives version as primary.

The omission of Shannon45 from the formal collection appears to have been an oversight resulting from insufficiently careful study of the report. To be honest, the reason I did virtually nothing with it for thirty years was that when I looked at it, it appeared to me to be the '48 paper, fleshed out with examples and tutorial material about particular cipher systems. Only later did I look carefully enough to note the much of the tutorial material was the development of information theory. I suspect that Sloane and Wyner, under the burden of editing the volume of Shannon's work, fell into much the same trap.

When I asked Neil Sloane why the '45 report had been omitted, he replied that "The paper we did include is identical to the 1945 memo, apart from the title." When I pointed out that "The '45 report has twice as many words and covers a variety of things not in the '49 paper, some of which are in the '48 paper." he responded: "When we were assembling his collected papers, we considered that report, and we decided that everything in it was included in the '48 and '49 papers." Although this it is not entirely true, it is close; most of what is in the '45 report is in the later papers. It does not, however, seem a sound reason for omitting the first version of some of Shannon's most important work. The '45 report is three years earlier and shows the motivation for his development of information theory.

It is entirely possible that there are other copies of Shannon's '45 memorandum still in existence. A logical starting place would be in the papers of the various people listed in the distribution list. Evidence of which version had wider distribution would suggest which was final. We would be grateful if anyone who

has or finds a fresh copy would let us know. Likewise, we would be grateful to know of any errors observed in this one.

We should mention that there is one significant error that we have left in the text. At the end of the second to last paragraph on page five we read: "In simple substitution with random key on English $|K|$ is $\log_{10} 26!$ or about 20 and D is about .7 for English. Thus unicity occurs at about 30 letters." The correct value for $\log_{10} 26!$ is about 27.

We are now fortunate to have the ancestral Shannon paper on information information theory publicly available in a clear legible edition for the use and enjoyment of cryptographers everywhere.