Further Information on Miller's 1882 One-Time Pad[*]

Steven M. Bellovin
Department of Computer Science
Columbia University
https://www.cs.columbia.edu/~smb

CUCS-011-16
November 27, 2016

Abstract

New information has been discovered about Frank Miller's 1882 one-time pad. These documents explain Miller's threat model and show that he had a reasonably deep understanding of the problem; they also suggest that his scheme was used more than had been supposed.

1 Introduction

Several years ago, I published a paper [4] describing how the one-time pad was invented by Sacramento banker Frank Miller in 1882 [11], more than 35 years before its reinvention by Gilbert Vernam and Joseph O. Mauborgne [5, 9]. Since then, I have discovered several new sources of information about Miller's work. Most important, I have found Miller's own explanation for why he created the scheme. In addition, some of the information casts doubts on my conclusion that his work sank into instant obscurity.

2 Miller's Explanation

The most interesting new discovery is Miller's own explanation of the theory behind his code book, published several months before the book itself [10]. This essay sets forth the basic principles of operation of the scheme:

> To secure secrecy the sender should have had prepared and filed with his correspondent a list of irregular numbers, such as 483, 281, etc., etc., which numbers should neither be repeated in any regular order, nor should the differences between them be in any

[*] Note: this is a preprint copy of a paper that has been submitted for publication. The final form may vary significantly.

> regular ratio. Furthermore, when one of these numbers has been used it should always be canceled for further use, so that the list may always show what number is next in order for use.

Miller had two threat models in mind. First, he was concerned about book-breaking based on probable plaintext derived from parallel telegrams:

> If the possessor shall desire to telegraph, “Pay John Jones one hundred dollars,” he can take for each of these words, except the proper nouns, the corresponding cipher word. If the [telegraph] operator has to send at the same moment an English dispatch from Smith to Jones, stating that an order is on the way to pay him one hundred dollars, then is the secrecy of the cryptogram dissolved, and furthermore the operator is ready to examine the next telegram in cipher and endeavor to ascertain if the same cipher word is used for “Pay,” as in the first telegram. Proceeding thus, day after day, he can detect word after word unless the owner of the code shall change his system often.

Miller was also very concerned about the integrity of messages, and in fact anticipated active attacks:

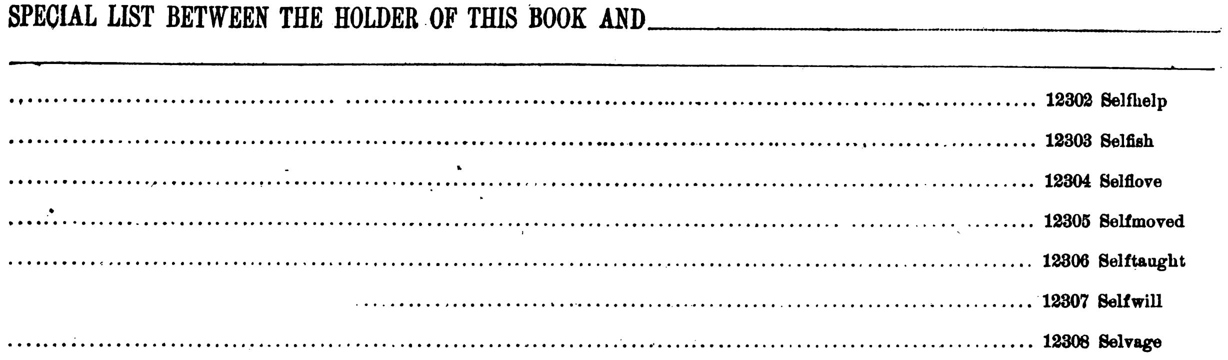
> Furthermore, such telegrams should contain passwords to satisfy the receiver that some shrewd villain has not cut the line just out of town and woven the ends into his surreptitious Morse instrument, thus enabling him to issue forged telegrams and direct payments to a confederate.

It also seems that he felt that telegraph operators were the most likely attackers, and that some such attacks had in fact occurred: “Great credit is due to the [telegraph] operators for their faithfulness. Few frauds have been committed by them, although they can easily decipher such ciphers by a little investigation.”

Miller seemed to anticipate Friedman’s argument for the absolute security of the one-time pad compared with how a conventional cipher’s cryptanalysis can be confirmed [5, 8]: “a tentative decipherment of a few letters in one message can be proved or disproved by the application of the resultant key letters to the homologous letters of the other cipher message.” Miller’s take was similar: “The detection of one telegram by means of parallel English messages will give no clue to past or future messages” and “no dispatch shall give a clue to any other, although all the dispatches may contain the same words.”

Finally, Miller understood the point-to-point nature of communications using a one-time pad: “We sill suppose that ten bankers have each such a book, and that one has sent to each of the remaining nine separate lists of ‘shift numbers,’ and that the others have followed the same plan.”

Interestingly, his article says nothing about randomness. He clearly understood the need for non-repetition, because repetition is at the heart of the attack he describes. However, he gives us no clue why he wrote that the “numbers should neither be repeated in regular order, nor should the differences between them be in any regular ratio.”



SPECIAL LIST BETWEEN THE HOLDER OF THIS BOOK AND ____________

12302 Selfhelp

12303 Selfish

12304 Selflove

12305 Selfmoved

12306 Selftaught

12307 Selfwill

12308 Selvage

Figure 1: The extra-word list in the Library of Congress and University of Chicago library's copies of Miller's book.

3 Obscurity?

In [4], I argued that it was unlikely that Miller's system had ever seen any real use. Issues included the lack of any provision for indicators as well as his clumsy rekeying instructions; in addition, C.F. Crocker, the vice president of the Southern Pacific Railroad, was listed as a holder of Miller's code but used a different and inferior confidentiality code [3]. A comparison of two of the three known extant copies of Miller's codebook casts doubt on my conclusion.[1]

Most telegraph codebooks had a list of blank extra words; Miller's was no exception, but his was unusual (Figure 1). I wrote:

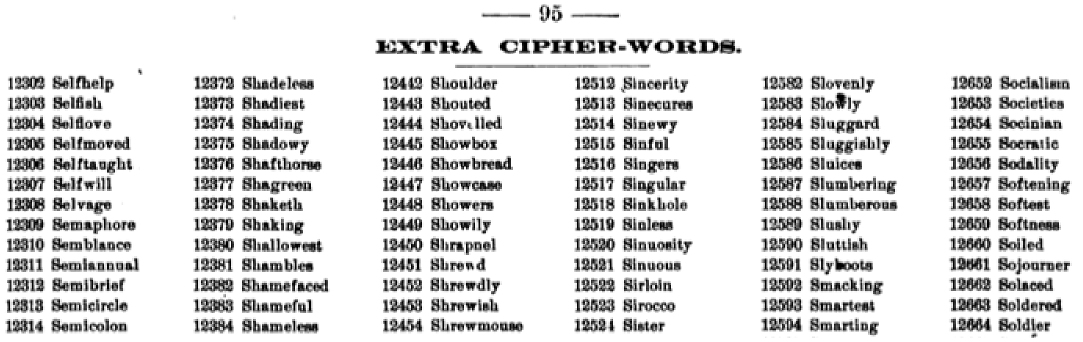
> Miller, by contrast, provided about 20 pages of pair-wise lists of extra words; each such page was intended for correspondence with a different individual. This clearly shows his orientation towards point-to-point communication, rather than one-to-many or many-to-many. Any realistic use of one-time pads would indeed require point-to-point messaging.

However, the New York Public Library's copy is different: it has a conventional list of extra words (Figure 2). Clearly, there were two different printings or editions; this is unlikely to have happened for a book that was never used.

It is uncertain which version is older, though it seems likely that the Library of Congress has the original one because of the requirements of the copyright law of the time: new works were required to be deposited within 10 days of publication [7]. This copy has a copyright stamp dated September 12, 1882, consistent with the reviews published a month later. Furthermore, the NYPL copy has an extra list of recipients. Miller would not have wanted to delete a list like that; if nothing else, it was good advertising, as well as more assurance of the utility of the codebook.

On the other hand, it is hard to understand why the separate point-to-point lists of extra words would have been deleted in favor of a single list. Possibly, customers did not understand the point, though even that would argue that there was a customer base.

[1] The University of Chicago copy is in poor condition and I have not seen it. However, inquiries via the librarian suggest that it is similar to the Library of Congress copy; in particular, it does not have the extra list of recipients.



— 95 —

EXTRA CIPHER-WORDS.

12302	Selfhelp	12372	Shadeless	12442	Shoulder	12512	Sincerity	12582	Slovenly	12652	Socialism
12303	Selfish	12373	Shadiest	12443	Shouted	12513	Sinecures	12583	Slowly	12653	Societies
12304	Selflove	12374	Shading	12444	Shovelled	12514	Sinewy	12584	Sluggard	12654	Socinian
12305	Selfmoved	12375	Shadowy	12445	Showbox	12515	Sinful	12585	Sluggishly	12655	Socratic
12306	Selftaught	12376	Shafthorse	12446	Showbread	12516	Singers	12586	Sluices	12656	Sodality
12307	Selfwill	12377	Shagreen	12447	Showcase	12517	Singular	12587	Slumbering	12657	Softening
12308	Selvage	12378	Shaketh	12448	Showers	12518	Sinkhole	12588	Slumberous	12658	Softest
12309	Semaphore	12379	Shaking	12449	Showily	12519	Sinless	12589	Slushy	12659	Softness
12310	Semblance	12380	Shallowest	12450	Shrapnel	12520	Sinuosity	12590	Sluttish	12660	Soiled
12311	Semiannual	12381	Shambles	12451	Shrewd	12521	Sinuous	12591	Slyboots	12661	Sojourner
12312	Semibrief	12382	Shamefaced	12452	Shrewdly	12522	Sirloin	12592	Smacking	12662	Solaced
12313	Semicircle	12383	Shameful	12453	Shrewish	12523	Sirocco	12593	Smartest	12663	Soldered
12314	Semicolon	12384	Shameless	12454	Shrewmouse	12524	Sister	12594	Smarting	12664	Soldier

Figure 2: The extra-word list in the New York Public Library's copy of Miller's book.

At least two publications [1, 2] carried reviews of Miller's book. The *New-York Tribune*'s review quite rightly focused on the security aspects of his code:

> He has devised therefore a scheme in which the meaning of the arbitraries *[sic]* is continually shifting. There is nothing new in that; but he has improved upon his predecessors by contriving not merely that the meaning of the ciphers shall be variable, but that the key for translation shall also vary with every word. A fortunate guess at the meaning of part of a dispatch would give no clew, in this system, to the meaning of the rest, nor any help in reading other dispatches between the same persons.

The review spoke of the additives as "chosen at random," even though Miller himself never used that word.

The article went on to say, "Obviously it is impossible to translate this cipher without the list of shift-numbers [additives], and so long as that is kept safe the system *seems* to afford absolute security" (emphasis added). It concluded, "We have examined a great many codes and this is the safest with which we are acquainted." It is fair to doubt that a general newspaper's telegraph code reviewer was an expert cryptanalyst (and the level of cryptologic knowledge in the U.S. at the time was quite pitiful); still, the reviewer did grasp the essence of Miller's scheme's confidentiality properties.

The other known review, in the *Banker's Magazine and Statistical Register*, is shorter and focuses somewhat more on the "test words"—authentication codes—in Miller's scheme. It does speak of security: "a feature which *seems* to render impossible forgery, or deciphering by any one except the correspondent holding the key" (emphasis added). It did not explain the cryptography as clearly as did [2], but instead pointed to Miller's own article [10].

4 Conclusions

If Miller's codebook was more secure than anything else and wasn't ignored, why did it fade into obscurity? Why was the concept of random, never-reused additives lost for more than 35 years? There are three reasons.

First, as Kahn notes, codebooks succeeded or fell based on both extrinsic and intrinsic factors [9]: “The extrinsic factor is the salesmanship of the compiler, and this often outweighs everything else.” Miller was a banker, not a telegrapher or a code compiler; if he didn’t actively market his system, it was bound to be replaced by something else. Second, it arguably wasn’t a very good codebook; there are too many individual words and too few phrases. Miller himself wrote of “the system of ‘packing’ long sentences each into one representative word, so as to save expense in ‘cabling’” [10]. In an era when codebooks were important enough that general circulation newspapers like the *New-York Tribune* deemed them worthy of book reviews (and this review was in a column along with a number of books aimed at general readers), there would have been many other choices.

Third, though, one-time pads are very hard to use in practice. Absent a serious threat model—and absent a clientele that had sufficient knowledge of cryptology to understand its advantages—the one-time pad was probably a disadvantage. There was probably a cryptologic problem—Miller wrote that the “deciphering [cryptanalysis] of important messages has been common, and sometimes has produced serious results. Among businessmen the topic is of daily importance”—but in the many-to-many world of bank transactions, one-time pads were probably the wrong solution. Governments could have used it, but the U.S. government was notoriously bad at cryptology.

The new information is interesting for the light it sheds on what Miller knew, and on whether or not his codebook was actually used. His understanding of the threats and of his system’s security were deeper than had been supposed.

Miller’s threat model goes a long way to explaining his desire for non-repetition and hence non-reuse. Additives were not new; they go back to the very first telegraph codebook [13]. However, the combination of reuse of an additive, repeated use of a codeword, and the probable plaintext he assumed would be available did add up to a threat, one that was exacerbated because the commercial codebooks of the day were generally one-part. Still, Miller took an unprecedented step beyond the common practice of using a repeating sequence of numbers, a notion that was proposed quite early [13]. The newly discovered documents do not shed any light on what led him to this insight.

Miller’s use of modular arithmetic is less surprising. Anyone trying to use additives in the real world would encounter the question of how to encrypt the last few entries; wrapping around is the obvious answer (though [13] had a more convoluted scheme that would have led to ambiguous decryptions). For example, Slater’s 1870 confidentiality code [12] said, “where the result exceeds 25000 [the size of his codeword space], deduct that number, or, in other words, commence the alphabet again.” Bloomer, in 1874, also realized this, though his solution was less clearly expressed [6]:

> To put the above sentences into a cipher word, it may be done by the sender adding, and the receiver deducting, or by the sender deducting, and the receiver adding; that is, counting a certain number of words forward or backward. . . To deduct from the first part, the last numbers will furnish the balance, and to add to the end, the first words will be the continuation.

The motivation for the randomness requirement remains quite unknown. It is a subtle point; the need for it is clearest to cryptanalysts, and there is as yet no evidence for Miller having any background in the subject. That said, this plus his concern about probable plaintext attacks does suggest some familiarity with cryptanalysis, either from his military background or perhaps from his conversations with the Sacramento Wells, Fargo agent and their concerns about rogue telegraph operators [4].

Finally, I also have no new information on whether or not Miller's ideas reached Mauborgne, perhaps via Parker Hitt. From a historical perspective, this is probably the most important question.

References

1. Book notices. *The Bankers' Magazine and Statistical Register*, 37(4):306, October 1882. Book review. URL: `http://search.proquest.com/docview/124425916`.

2. New publications. *New-York Tribune*, page 6, October 6, 1882. Book review. URL: `http://search.proquest.com/docview/573003445`.

3. *Official Cipher*. 1893? Apparent private cipher book, with numbered copies held by several California notables. Copy consulted is in the Huntington Library.

4. Steven M. Bellovin. Frank Miller: Inventor of the one-time pad. *Cryptologia*, 35(3):203–222, July 2011. An earlier version is available as technical report CUCS-009-11. URL: `http://dx.doi.org/10.1080/01611194.2011.583711`.

5. Steven M. Bellovin. Vernam, Mauborgne, and Friedman: The one-time pad and the index of coincidence. In Peter Y. A. Ryan, David Naccache, and Jean-Jacques Quisquater, editors, *The New Codebreakers: Essays Dedicated to David Kahn on the Occasion of His 85th Birthday*. Springer, 2016.

6. J. G. Bloomer. *Bloomer's Commercial Cryptograph: A Telegraph Code and Double Index—Holocryptic Cipher*. A. Roman & Co., San Francisco, 1874. URL: `http://books.google.com/books?id=90UKAAAAIAAJ`.

7. Robert Brauneis. Private communication, November 24, 2016.

8. William F. Friedman. Differential primary keys in cryptography. In *Item 1056, William F. Frieman Collection, George Marshall Foundation Library, Lexingon, VA*. Although the actual manuscript was created no earlier than 1924, it contains a typed headnote saying that it was based on materials prepared at Riverbank in 1920.

9. David Kahn. *The Codebreakers*. Macmillan, New York, 1967.

10. Frank Miller. Cipher telegrams. *The Bankers' Magazine and Statistical Register*, 36(9):662, March 1882. URL: `http://search.proquest.com/docview/124403296/`.

11. Frank Miller. *Telegraphic code to Insure Privacy and Secrecy in the Transmission of Telegrams*. Charles M. Cornwell, New York, 1882. Google graciously scanned this book at my request. URL: `http://books.google.com/books?id=tT9WAAAAYAAJ&pg=PA1#v=onepage&q&f=false`.

12. Robert Slater. *Telegraphic Code, to Ensure Secresy in the Transmission of Telegrams*. W.R. Gray, London, first edition, 1870. URL: `http://books.google.com/books?id=MJYBAAAAQAAJ`.

13. Francis O.J. Smith. *The Secret Corresponding Vocabulary, Adapted for use to Morse's Electro-Magnetic Telegraph: and Also in Conducting Written Correspondence, Transmitted by the Mails, or Otherwise*. Thurston, Ilsley & Co., Portland, ME, 1845. URL: `http://books.google.com/books?id=Z45clCxsF7EC`.