

Re: Confirmation of Shannon s Mistake about Perfect Secrecy of One-time-pad

Source: <http://sci.tech-archive.net/Archive/sci.math/2007-10/msg04050.html>

- *From:* hagman <google@xxxxxxxxxxxxxxx>
 - *Date:* Mon, 22 Oct 2007 12:53:07 -0700
-

On 22 Okt., 06:22, wangyong <hell...@xxxxxxx> wrote:

Confirmation of Shannon s Mistake about Perfect Secrecy of One-time-pad

Yong WANG

School of Computer and Control, GuiLin University Of Electronic Technology ,Guilin, 541004, Guangxi Province, China
hell...@xxxxxxx

Abstract: This paper analyzes Shannon s proof that one-time-pad is perfectly secure and discusses all the possibilities, and then confirms which of them is Shannon s notion. The confirmed notion is found to be wrong for its neglect of the prior probability and the absolutely irreconcilable conditions. Therefore, one-time-pad is not perfectly secure.

Keywords: one-time system, cryptography, perfect secrecy, information theory, probability

1. Introduction

Shannon put forward the concept of perfect secrecy and proved that one-time-pad (one-time system, OTP) was perfectly secure [1, 2]. For a long time, OPT has been thought to be unbroken and is still used to encrypt high security information. In literature [3], example was given to prove that OPT was not perfectly secure. In literature [4], detailed analyses about the mistake of Shannon s proof were given.. It was proven that more requirements are needed for OTP to be perfectly secure and homophonic substitution could make OTP approach perfect secrecy [5]. Literature [6] analyzed the problem and gave ways to disguise the length of plaintext. In literature [7], the cryptanalysis method based on probability was presented, and the method was used to attack one-time-pad. Literature [4] considered an especially understanding following which OTP could be thought perfectly secure if some added conditions were satisfied. In this paper, I will confirm that the especially understanding is not Shannon s notion and analyze the source of his mistake.

2. Counterexample of Shannon s Conclusion

For the moment, we do not consider the inaccessible limitation that the length of any plaintext must be the same as the length of ciphertext in OTP. The following discussion supposes the all the

plaintexts, ciphertexts and keys are the same in length.

We give a simple example of OTP to discuss the problem, plaintext space is $M = \{0,1\}$, ciphertext space is $C = \{0,1\}$ and key space is $K = \{0,1\}$. According to the information that cryptanalysts got beforehand, they can get the prior probability of plaintext as $P(M=0) = 0.9$ and $P(M=1) = 0.1$. Later the ciphertext $C=0$ is intercepted. When only considering $C=0$ and the cryptosystem (regardless of the prior probability of plaintext), we can deduce that the plaintexts are equally likely, for there is a one-to-one correspondence between all the plaintexts and keys for $C=0$. The prior probabilities of plaintexts are seldom the same, so the two probability distributions of plaintexts gained from different conditions are conflicting. Then the compromise of the two probability distributions is indispensable. The compromised posterior probability of the plaintext would be between the two corresponding probabilities of the two sectional conditions. When $C=0$ is intercepted, the posterior probability $P(M=0)$ is between 0.9 and 0.5, and $P(M=1)$ is between 0.1 and 0.5. The compromised posterior probability of the plaintext isn't equal to the prior probability, so OTP is not perfectly secure.

3. Mistake Confirmation of Shannon's Proof

Due to the mapping of M , K and C , the probabilities of M , K and C are complicatedly interactional. For the above example, the probability of plaintext changes when the ciphertext is fixed, even though the ciphertext is unknown.

When only considering the fixed ciphertext and the equiprobability of key, we can gain that plaintexts are equally likely for there is a one-to-one correspondence between all the plaintexts and keys for fixed ciphertext. There is conflict between the prior probability and the uniformly distributed probability gained above.

In order to understand the inconsistency of probability in the example and the need for fusion of the probabilities in this case, we adopt the combinations of different conditions for the following deduction to analyze the existence of probability conflict.

For our simple example about OTP, when considering the condition that ciphertext is 0, the probability of ciphertext being 0 is 1, and the probability of ciphertext being 1 is 0. But according to the prior probability distribution of plaintexts given and uniformly distributed keys, we can easily find that ciphertext is uniformly distributed, that is to say, all ciphertext are equally likely. We can see the two probability distributions of ciphertext in different conditions are conflictive.

When only considering that the intercepted ciphertext is 0 and prior probability of plaintext being 0 we call $P(M=0)$ is 0.9, and prior probability of plaintext being 1 we call $P(M=1)$ is 0.1, the probability of key being 0 we call $P(K=0)$ is 0.9, and the probability of key being 1 we call $P(K=1)$ is 0.1 because there is a one-to-one correspondence between all the plaintexts and keys. However, according to the requirement of OTP, all the keys are equally likely, so conflict of the probabilities occurs as before.

Such conflicts show that under different conditions we may draw inconsistent probabilities, so it needs to fuse and compromise. The

probabilities obtained by the different combinations of unilateral conditions are inconsistent. That is to say, the conditions in OPT can not coexist. When all the conditions are considered, some of the conditions must change, so it is not proper to use these conditions when computing the final posterior probability. It likes four irregular feet of a same table. There is always one foot that is turnup when the table is on the horizontal ground. If the four feet should touch the horizontal ground at the same time, distortion would happen. In literature [7], formula was presented to fuse the inconsistent probabilities.

Shannon did not realize that the conditions were impossible to coexist. When taking them into the formula, there must be mistake for the conditions cannot coexist and the probabilities have changed when all the conditions are considered at the same time.

For the conditions in the example are very complex, and some are connotative, it is essential to list them and analyze the impact of the conditions on the probability. Literature [4] considered an especially understanding following which OTP could be thought perfectly secure if some added conditions were satisfied and analyzed that was unlikely to be Shannon's view. This paper analyzes the problem in detail and confirms the result that the especially understanding is not Shannon's view using the information gained from Shannon's proof.

As different conditions can gain different probability distribution, we list the conditions those impact on the probability distribution of plaintext and the corresponding probabilities of plaintext when only considering some of the conditions.

A Considering the information that cryptanalysts got beforehand, we can get $P_1(M=0)=0.9$, $P_1(M=1)=0.1$

B Considering the cryptosystem (including that the keys are equally likely) and unknown but fixed ciphertext(C can be 0 or 1, but not random variable), we can get $P_2(M=0)=0.5$ and $P_2(M=1)=0.5$ for there is a one-to-one correspondence between all the plaintexts and keys for fixed ciphertext.

C Considering the cryptosystem and known ciphertext $C=0$, we can get $P_3(M=0)=0.5$ and $P_3(M=1)=0.5$ for there is a one-to-one correspondence between all the plaintexts and keys for fixed ciphertext.

D Considering the information that cryptanalysts got beforehand and the cryptosystem, we can get $P_4(M=0)=0.9$, $P_4(M=1)=0.1$ for the cryptosystem does not impact on the probability of plaintext.

E Considering the information that cryptanalysts got beforehand, the cryptosystem and unknown but fixed ciphertext, we can get that $P_5(M=0)$ is between 0.9 and 0.5 and $P_5(M=1)$ is between 0.1 and 0.5 after compromise.

F Considering the information that cryptanalysts got beforehand, the cryptosystem and ciphertext $C=0$, we can get that $P_6(M=0)$ is between 0.9 and 0.5, $P_6(M=1)$ is between 0.1 and 0.5 after compromise.

Posterior probability is known from conditional probability, but there is still precondition or information to get prior probability. Suppose that we have no idea of an event, we can't know how many possible

Re: Confirmation of Shannon's Mistake about Perfect Secrecy of One-time-pad

random values there are, not to mention the corresponding probabilities of the possible values. Therefore, the prior probability is based on the known conditions and it is also a conditional probability. Shannon did not confirm what the precondition for the prior probability is and did not define prior probability definitely. Considering the possibility of wrong understanding, prior probability may be one of (1), (4) and (5). Posterior probability may be one of (3) and (6). If we take prior probability as that in (5) and take posterior probability as that in (6). The two probabilities may be the same. Suppose that is Shannon's view, he should have the concept of information fusion and could distinguish the conditions, but he never did so. What's more, there was no algorithm to fuse the probabilities at that time. Shannon always thought the fixedness of probability to be perfectly secure. In the especially supposed case, the initial probability in (1) is still changed and translated to the probability in (5), the change should be thought to be not secure by Shannon's view.

Strictly speaking, the posterior probability should be that in (6), but there is another understanding that the posterior probability is that in (3) for it is not distinctly defined. In (3) we can easily get the probability and do not have to explain the conflictive probabilities, but the probability in (6) is hard to understand and compute because the probabilities in different partial conditions are inconsistent and compromise is needed. We can find that Shannon did not realize the problem of compromise from his proof, so Shannon may take the posterior probability as that in (3). What's more, we can confirm that Shannon thought of the posterior probability as that in (3) from his following proof [2]:

It is possible to obtain perfect secrecy with only this number of keys, as one shows by the following example: Let the M_i be numbered 1 to n and the E_i the same, and using n keys let

$$T_i M_j = E_s$$

where $s = i + j \pmod{n}$. In this case we see that $P_E(M) = 1/n = P(E)$ and we have perfect secrecy.

$P(E)$ = probability of obtaining cryptogram E from any cause.

$P_E(M)$ = a posteriori probability of message M if cryptogram E is intercepted.

$P(M)$ = a priori probability of message M

Shannon got the result $P_E(M) = 1/n = P(E)$, that is to say, the plaintexts are equally likely when the cryptogram E is intercepted. That is just the case of (3). But it is wrong for the prior probability of plaintext is not considered; otherwise the plaintexts are not equally likely. Shannon seemed to deem the result was very easy to get for Shannon got the result without strict proof in detail. If he considered the case in (6) but not that in (3), the problem would be very complex.

We can confirm Shannon's mistake by using his result to get cockeyed result. Using Shannon's result that the given example is perfectly secure, we can get $P_E(M) = P(M)$, as Shannon got $P_E(M) = 1/n$, so we can get $P(M) = 1/n$. But that is wrong for plaintexts are seldom equally likely.

Re: Confirmation of Shannon s Mistake about Perfect Secrecy of One–time–pad

In summary, Shannon is wrong and one–time–pad is not perfectly secure.

The especially understanding is not Shannon s view.

Shannon first proved that OTP was perfectly secure, but his proof is very simple. Detailed proofs about perfect secrecy of OTP were given by later scholars who used Shannon s proof for reference. There are many proofs about perfect secrecy of OTP. Some proofs directly made the conclusion that all plaintexts were equally likely. But it is obviously wrong, for the prior probabilities of all plaintexts are seldom equally likely. Other proofs are generally identical with minor differences. It is found that the latter proofs made similar mistakes.

4. Conclusion

The paper analyzes Shannon s proof, discusses all the possibilities and confirms which of them Shannon s notion is. The confirmed notion is found to be wrong for its neglect of the prior probability and the absolutely irreconcilable conditions. Though one–time–pad is not perfect secure, it still has good cryptographic property. We can take measures to improve its security. The above analyses not only confirm Shannon s mistake, but also fetch out the limitation of probability theory and information theory. In the two theories, the value of probability is deemed to be fixed, but in some cases, probability may be random variable in practice [8, 9].

Reference

- [1]. Bruce Schneier, Applied Cryptography Second Edition: protocols, algorithms, and source code in C[M],John Wiley & Sons, Inc, 1996
- [2]. C.E.Shannon, Communication Theory of Secrecy Systems [J], Bell System Technical journal, v.28, n.4, 1949, 656–715.
- [3]. Yong WANG, Security of One–time System and New Secure System [J], Netinfo Security, 2004, (7):41–43
- [4]. Yong WANG, Fanglai ZHU, Reconsideration of Perfect Secrecy, Computer Engineering, 2007, 33A9
- [5]. Yong WANG, Perfect Secrecy and Its Implement [J], Network & Computer Security, 2005(05)
- [6]. Yong WANG, Fanglai ZHU, Security Analysis of One–time System and Its Betterment, Journal of Sichuan University (Engineering Science Edition), 2007, supp. 39(5):222–225
- [7]. Yong WANG, Shengyuan Zhou, On Probability Attack, Information Security and Communications Privacy, 2007,(8)Ó910
- [8]. Yong WANG, On Relativity of Probability, www.paper.edu.cn, Aug, 27, 2007.
- [9]. Yong WANG, On the Perversion of information s Definition, presented at First National Conference on Social Information Science in 2007, Wuhan, China, 2007.

The Project Supported by Guangxi Science Foundation (0640171) and Modern Communication National Key Laboratory Foundation (No.

9140C1101050706)

Biography

Re: Confirmation of Shannon s Mistake about Perfect Secrecy of One-time-pad

If I understand you right, then you can get more characters right than by guessing from the encrypted message found below.

I use the following C code:

```
#include <stdio.h>
#include <stdlib.h>

int main(int argc, char* argv[] ) {
    srand( atoi(argv[1]) );
    int limit = atoi(argv[2]) * (RAND_MAX/100);
    for (int row = 0; row < 20; ++row) {
        for (int col=0; col < 50; ++col)
            printf( (rand()<limit)? "0" : "1" );
        printf("\n");
    }
}
```

to generate a 1000 character plaintext with $P(M=0) = 0.9$:

```
#!/foo (non-disclosed-random-seed) 90 > plaintext
```

and a onetime pad

```
#!/foo (non-disclosed-random-seed) 50 > onetimepad
```

Just to make sure I cannot change the data afterwards behind your back:

```
#cat plaintext onetimepad | md5sum
044fc613804c45291d6e557a1ef265da -
```

Applying the onetimepad to the plaintext (xor'ing them), I obtain

```
00001011001110101101100100010000001010101110000010
10100110010110101111100011011101110101010100101011
00011101101111111101111010111101000101111100101000
01000111100101010000000100000110110111100011010111
10110101010111001001100010001010000110100010100100
0111001111000110010011111011111000111100111100111
10000110000111010001000001100011011011000001110000
01100011001010011000000100010000111011111101101000
10101001111101011000100101110111110110000101010001
11010101011110110110000000011011111111111010101111
00101011011101111010000101011111111100000101010100
11011111110011101110011010111000110001011100111110
11011011110001000110101001111110011001101001011010
01100110001101000101110001011000110010110111011001
00111001010000000011101011110010111010011101000111
10111100011110110011111011001010111010111001000010
01110111111101111000101101110010010100001011101100
00111110011000111000000110011001001011111111111110
00001010010111110111110110000000011000011110001110
01111111011000101001011101101100101010110011000010
```

Your task is to guess the plaintext or rather to be significantly

Re: Confirmation of Shannon s Mistake about Perfect Secrecy of One-time-pad

better than guessing: Simply saying "0000...00" will get ~900 characters right; producing a new random text with $P(M=0)=0.9$ would get about $0.9*900$ 0's and about $0.1*100$ 1's correct, in total only ~820.

But can *your* theory produce a suggested decrypted text that coincides with the plaintext at significantly more than 900 characters? It seems to claim so.

If you can post a text that has 905 or more correct characters, I'll read your paper more thoroughly and become one of your supporters.

If you can get 910 or more characters right, you'll have totally shattered my gut feelings about one-time pads.

Will you try?

hagman

.