UDC 621.391.82

POTENTIALLY UNBREACABLE CIPHERING ON A HYBRID PHYSICAL-MATHEMATICAL LEVEL

A. G. Tyzhnenko, E. V. Ryeznik

Department of Mathematics, Kharkiv National Economic University
9-A Lenin Ave., Ukraine
e-mail:a.tyzhnenko@gmail.com

Ciphering based on stochastic comparison of signals in electromagnetic form or in numerical one is addressed. Due to such a comparison, one can identify a stochastic signal as some binary unit of information. That is, each bit of information can be transmitted as a stochastic signal with or without carrier. At a receiver, identification of bits (1s or 0s) is proposed to do by measuring the correlation coefficient between received signal and two replicas, one of which corresponds to 1s and the other corresponds to 0s. To increase security level, some normal noise is intentionally added to transmitted signals and signal samples are mixed with noise samples. These ones prevent decrypting both on physical and mathematical levels. Namely these issues are used in the work for constructing a potentially unbreakable scheme of communication in which the security problem is solved on combined physical and mathematical level.
KEY WORDS: bit, stochastic signals, safe communication, hard drive protection

INTRODUCTION

The safe communication problem is centuries old. Nowadays, it exhibits new features caused by impersonal electronic communications. For a gateway to the literature, one can see [1-3], for example. Because of these new features, the safe communication problem can be split into two parts:

- Type I problem – the safe data distribution;
- Type II problem – the computer's hard drive protection.

These problems are commonly considering on three levels:

- Mathematical level – encryption;
- Software level – that based on the Net and computer properties;
- Physical level – that based on physical principles.

To briefly outline the state of the art of safe communication problems, we consider some methods, the most interesting, as to our mind, from the wide variety of existed and potential ones.

The one-time-pad [1, 2, 3] is considering now as the most durable encrypting scheme. It solves the Type I problem. The most serious shortcoming is the key distribution problem, which causes the same security problem as that for data transmission. More than that, the key length must be exactly the same as that of the message that is not convenient. This method works on mathematical level.

The one-time-pad with quantum encrypting [3] is very interesting potential method, which is designing to solve the Type I problem on physical level. When this method will be completed, it seems to be a perfect encrypting method, which can completely solve the Type I problem.

The most widely used in practice encryption method is probably the public-key scheme [3]. It solves the Type I problem on sufficiently high level with the aid of mathematical principles. This method is not perfect, but gives very strong and convenient encryption that is sufficient, in main, for the common workers in the Internet.

One else very interesting encryption scheme based on physical principle, named as chaotic communication [4-7], is nowadays developing. This scheme is designing to perfectly solve the Type I problem. It seems to be the case after solving the synchronization signal safe transmission problem.

To solve both the Type I and Type II problems on software level the firewall schemes have been designed [1]. As any software scheme, a firewall does not provide perfect security. However, it can solve the safe communication problem on a high level sufficient for common use in the Net.

Summing, we can note that the safe communication problem is solved to date on some practically sufficient level for common utilization in the Net. What is remained unsolvable is totally safe communication problem needed for special cases. We propose in this work a new scheme that solves this problem on a combined physical and mathematical level. Totally secure communication may be obtained at the expense of bit rate decreasing in the data exchange process. In the core of the method proposed lies a physical principle according to which a would-be eavesdropper or intruder has limited opportunity for bit identification compared to that at a receiver. This is because the identification level for an eavesdropper is intentionally diminished by adding noise to signal bits and mixing signal samples with noise samples. Due to this one decrypting become impossible regardless of a computer system used for this purpose.

BITS CAMOUFLAGING AND IDENTIFICATION PROBLEMS

In the case of binary encoded information, each bit (1 or 0) is associated in our method with some definite information-bearing signal, which consists of the set of equidistant in time samples:

$$x = \{x_i\}_{i=1:N}, \tag{1}$$

where $x_i = x(t_i)$, are peak values spaced uniformly in time with discretization period Δ. That is, the set $\{t_i\}_{i=1:N}$ is also spaced equidistant in time. Thus, each bit of a message is supposed to be transmitted as the *N*-length signal of the type (1), and the number of samples, *N*, is unique for each communication line. Let all 1s are transmitted with the aid of any one but fixed *N*-length realization of pseudorandom signal,

$$x^{(1)} = \{x_i^{(1)}\}_{i=1:N}, \tag{2}$$

and all 0s - with the aid of any other fixed *N*-length realization from the same population,

$$x^{(0)} = \{x_i^{(0)}\}_{i=1:N}. \tag{3}$$

As the bit-bearing signals, we use in this work the Gauss normal distributed samples with population mean zero and standard deviation one of the type (2), (3). These signals are distributed from transmitter to receiver instead of 1s and 0s. While transmitting, original signals (2), (3) are altered due to presence of noise and boosting, and become as the following:

$$\tilde{x}^{(1)} = ax^{(1)} + z, \tag{4}$$

$$\tilde{x}^{(0)} = bx^{(0)} + z. \tag{5}$$

Here, $z = \{z_i\}_{i=1:N}$ is any realization of white Gaussian noise. Note that there are different realizations of noise *z* in (4) and (5) due to randomly affected noise environment. At the same time, the signals (2) and (3) are the same for all 1s and 0s correspondently. The amplitudes *a* and *b* may be not equal to one as in (2), (3) due to boosting, for instance. At a receiver, the incoming signals (4), (5) are compared with known replicas, which are exactly the same as the original signals (2), (3). As the comparison method, we use here the correlation coefficient between two signals, *x* and *y*, of the same length, although it is not the most efficient method. Such a comparison method is known as a robust and independent from multiplication of signals by constant value. The last issue is very important in communications because the received and original signals may have different amplitudes, *a* and *b* in (4), (5) compared to the ones in (2), (3).

That is, the identification of bits at a receiver is done by measuring correlation coefficients between incoming signals and replicas:

$$r_{11} = r(x^{(1)}, \tilde{x}^{(1)}),\ r_{00} = r(x^{(0)}, \tilde{x}^{(0)}) \tag{6}$$

(between 1s and between 0s correspondently), and also between 1s and 0s:

$$r_{10} = r(x^{(1)}, \tilde{x}^{(0)}),\ r_{01} = r(x^{(0)}, \tilde{x}^{(1)}). \tag{7}$$

The measurements are fulfilled at *N* equidistant points $\{t_i\}_{i=1:N}$, if we know exactly the beginning moment of the message. Otherwise, we have to measure the correlation coefficients at *N* equidistant points as well, but in this case, we have to do this 10 times, for example, on the interval between samples. Such a discretization is approximately optimal.

It is clear the coefficients (6) are close to one but the remains (7) are close to zero. To identify the bit 1, we define the threshold $k^{(1)}$ as $k^{(1)} = (\bar{r}_{11} + \bar{r}_{10})/2$. Analogously, to identify the bit 0, we define the threshold $k^{(0)}$ as $k^{(0)} = (\bar{r}_{00} + \bar{r}_{01})/2$. Here, the quantities $\bar{r}_{11}, \bar{r}_{10}, \bar{r}_{00}, \bar{r}_{01}$ are corresponding mean values of coefficients (6)-(7) and can be estimated with the aid of the Monte Carlo experiment for known noise level. Technically, the coefficients (6)-(7) are measured on *N*-points set $\{t_i\}_{i=1:N}$ for each simulation by discretization period Δ or by one tenth of this period. If the correlation coefficient, $r(x, x^{(1)})$, measured in real time exceeds the critical value $k^{(1)}$, then the *N*-length signal $\{x_i\}$ represents the bit 1 on some confidence level that can be estimated theoretically. Analogously, if measured $r(x, x^{(0)})$ exceeds the critical value $k^{(0)}$, then the *N*-length signal $\{x_i\}$ represents the bit 0 on the same, practically, confidence level.

The scheme of secure communication is based on above considerations. With the goal of increasing security level, we enlarge intentionally the level of noise *z*. Then, a cryptanalyst have to compare two signals with noise, namely, $x + z$ and $x + z'$, where z and z' are different noise realizations. Because of that, the maximum correlation coefficient, which a cryptanalyst can find during scanning the message, is the following: $r_{cr} = r(x + z, x + z')$. On the other hand, a receiver registers significantly larger correlation coefficient $r_{rec}^{(0,1)} = r(x^{(0,1)}, x^{(0,1)} + z)$ because replicas at a receiver do not contain noise. Here, *x* represents the part of a message used by a cryptanalyst for scanning the message with the goal of finding bits. If $x = x^{(1)}$, for example,

then cryptanalysts sees $r_{cr} = r(x^{(1)} + z, x^{(1)} + z')$, but receiver registers $r_{rec}^{(1)} = r(x^{(1)}, x^{(1)} + z)$, and $|r_{cr}| << |r_{rec}^{(1)}|$.

This issue is due to presence of artificial noise added to bit-bearing signal transmitted. This one permits to identify bits at a receiver on darn small Type I error level, which depends on both the SNR and bit-bearing signal length. On the other hand, a cryptanalysts can identify bits only on sufficiently higher error level. As it will be further shown, namely this physical issue prevents a would-be cryptanalysts from finding bits with fidelity.

However, only this issue cannot prevent completely the bits identification by a cryptanalyst. This is because we cannot add so strong noise to bit-bearing signals that the Type I error of bits identification for an eavesdropper will be large enough to prevent totally such identification. If we will do that, the bit error rate at a receiver will not be sufficiently small. To improve the situation, we use 2*N*-length packet to transmit each *N*-length bit-bearing signal. The *N*, from each 2*N* randomly distributed samples of each packet, consist of bit-bearing samples, $\{x_i + z_i\}_{i=1:N}$, placed randomly among all other samples. Other *N* samples are white Gaussian noise ones, $\{g_m\}_{m=1:N}$. Full packet, which represents a bit, can be then represented mathematically as $\{x_i + z_i\} \bigcup \{g_m\}$. The distribution feature of bit-bearing signal samples, $\{x_i + z_i\}_{i=1:N}$, among 2*N* places in the packet is unique and known at transmitter and receiver only. Such camouflaging scheme can be so optimized that it can become totally unbreakable regardless of cryptanalysis method employed and computer possibilities. To prove this, consider the bit identification process at a receiver and compare it to that which can be used by a cryptanalyst. At a receiver, the presence of additional noise $\{g_m\}$ does not change the identification process because receiver knows the bit-bearing samples distribution feature inside the packet. An eavesdropper, on the other hand, does not know this one. To find bits, he has to investigate the message's structure. However, as is clear from above, this structure is statistically uniform. Indeed, both bit-bearing signals (4), (5) are some fixed realizations from the same normal population. To prevent repeating of these signals in the message, we add to them random realizations of white Gaussian noise with signal to noise ratio about one. Hence, the bit-bearing signals (4), (5) have the same statistical features and non-repeating structures; the noise $\{g_m\}$ is random realization from the same population. These ones prevent to find bits with the aid of finding the structure repeated. Because of that, the only method for a cryptanalyst to find bits is the same that uses a receiver while registering bits, viz., the correlation coefficient measuring.

To this end, a bit-finder has to take a part of the message as a base *2S*-length packet. In this base packet approximately half samples belong to signal samples. An analyst may suppose that such signal samples are exactly *S*. Picking up randomly *S* samples, supposedly bit-bearing samples, from the base packet an analyst has to scan step by step the rest of the message finding the sharp increasing in the correlation coefficient. Phenomenologically, this is the only way to find bits. However, this way can produce no result if the base signal does not contain some bit-bearing samples. Hence, the length of a base signal, *2S*, must be large enough, but not larger than the length of signal packet to avoid the obliged overlapping case. So far as the length of signal packet is not known, it is not possible to choose correctly the length of the base packet for scanning. Let us suppose that bit-finder knows the range of possible sizes of signal packets commonly used. Let it be from 80 to 200 samples. That one forces *S* to range from 40 to 100 while finding bits.

If a bit-finder wants to have *S* bit-bearing samples in the base signal, he should take the base packet length in any case more than 2*S*. Further, we take it 2*S* for simplicity while estimating the complexity of bit-finder's efforts, regarding it will be an underestimation. That is, let us consider that from 2*S* randomly chosen samples (base packet), *S* samples belong to the bit-bearing signal, but an analyst does not know which of them specifically. In this case, he has to try various combinations from 2*S* samples by *S* ones. Such combinations will be equal to C_{2S}^{S} regarding the order of signal samples is not changed from bit to bit. With each such a combination in hand, an analyst should evaluate the correlation coefficients between the base *S*-length signal and each *S*-length part outside the chosen 2S-length base packet approximately $10^7 C_{2S}^{S}$ times while scanning step by step the whole message finding 1s or 0s. If the chosen base packet belongs entirely to a signal packet, a bit-finder can, potentially, identify one from two bits:

$$0 \text{ or } 1. \tag{8}$$

Note that the bit-finding process is complicated if the base packet overlaps two adjacent signal packets. In this case analyst will find one of four bit combinations:

$$0\text{-}0,\ 0\text{-}1,\ 1\text{-}0,\ 1\text{-}1. \tag{9}$$

The identification process can be done on the statistical level only. To this end, mean correlation coefficient ($\bar{r}$) and its STD (σ) has be computed for each scanning procedure. Each case which gives correlation coefficient larger than 3σ level:

$$r > \bar{r} + 3\sigma, \tag{10}$$

must be verified as potentially signal samples. To this end, one has to determine the signal packet length (2*N*) with most accuracy as possible. For this one, an analyst has to add more samples, at least 2*S*, to both sides of base packet. From these added samples one has to choose additional samples finding those belonged to signal

samples. Each time one has to compute the correlation coefficient with samples correspondent to maximal r (10). The signal packet length ($2N$) may be theoretically obtained finding maximal correlation coefficient for all combinations of added samples. Nevertheless, $2N$ can not be derived exactly because of presence noise in signal samples because this one forces random behavior of correlation coefficient while adding samples. Deriving of approximate signal packet length for one maximum only requires for $\sum_{i=1}^{S} C_{4S}^{i}$ computations of correlation coefficient. Accounting for S changing from 40 to 100 and semantic analysis needed for separation cases (8) and (9), the identification process needs for approximately

$$\nu = 10^7 \sum_{S=40}^{100} (10^7 C_{2S}^{S} + M_S \sum_{i=1}^{S} C_{4S}^{S}) \tag{11}$$

computations of correlation coefficient. Here, M_S is the mean number of maximal correlation coefficients satisfied the condition (10). Number ν is enormously large and because such finding process is not realistic regardless of a computer system used for this purpose.

Such a conclusion has a phenomenological meaning. Inverse number of (11) estimates approximately a theoretical probability of finding the needed distribution feature of all N bit-bearing samples in the packet. So small probability of signal bits identification for an analyst is caused by relatively high identification error both due to noise added to signal samples and additional noise samples presented in signal packet.

This one produces the second crucial point, which prevents bits identification with fidelity. Indeed, because of huge amount of trails in the finding process, the error level of bits identification must be darn near to zero, in order to identify bits with fidelity by an analyst. However, due to the added noise in each packet, the Type I error of bit identification by a cryptanalyst can be maintained on relatively high level, namely, about of (1-10)%. This issue completely prevents a cryptanalyst from bit identification on some acceptable level because of among huge number of trails there will be enormously large cases of error identification.

There exists, however, another one potential opportunity for a cryptanalyst to find bits. He can hypothetically use the whole $2N$-length signal as the base one without picking up the signal samples. Such signal contains both all bit-bearing samples with noise, $\{x_i^{(0,1)} + z\}$, and strong noise ones, $\{g_m\}$. However, due to existence of strong noise samples, $\{g_m\}$, the correlation coefficient between the base and message signals is not distinguished on any sufficiently significant level from background correlation in this case. This point is demonstrated further.

It is very important to note that decryption scenario does not depend realistically on computer possibilities. Such situation is, really, a result of following causes:

- The presence of noise in bit-bearing signals;
- The random distribution of bit-bearing samples among additional noise samples in bit-bearing packets.

This issue depends only on the length N of bit-bearing signals and signal to noise ratio in both the bit-bearing signal and other samples in the packet. Due to the first cause, a receiver's correlation coefficients for bits are significantly larger than those for a cryptanalyst. Due to the second cause, to find bits a cryptanalyst has to try huge number of combination. This one prevents to identify bits with fidelity because the identification error (α) is sufficiently large, and erroneous identification could be realized too many times in the finding process.

Regarding to the above, we have to note that the method supposed requires unconventional data acquisition scheme in which magnitude of incoming signal samples are measured, decrypted, and represented as bits in modem. After decrypting, modem transmits bits to computer. Such a scheme prevents from transmitting some illegal bit-represented information inside the computer area. The latter one prevents completely a computer from any intruding. To neutralize possible stealing of information about signal samples, distribution features of samples inside signal packet and its length, all these things can be changed each time immediately in a new message. New features can be transmitted in the frontal part of communication.

NUMERICAL SIMULATIONS OF THE SAS

Now, we demonstrate some possibilities and merits of the above considered stochastic communication scheme (SAS). At first, we should note that, as known, the correlation coefficients are normally distributed. This one permits to estimate with fidelity the identification error (the Type I error, α) at a receiver and for a would-be cryptanalyst. Because of principle importance of this issue, this point has been justified in our research with the aid of the Kolmogorov-Smirnov criterion.

From the bit-rate point of view, the length of bit-bearing packets should be as shorter as possible. On the other hand, the sorter is the packet the harder is to camouflage bits completely in noise. The compromise can be achieved in each specific case separately. Consider the more or less optimal case when the bit-bearing signal consists of 50 samples (N = 50). In this case, it is rational to chose the signal to noise ratio in bit-bearing signal, SNR = 0.8, and the signal to noise ratio in the rest part of the bit-bearing packet, $SNRN$ = 0.2. To calculate

statistical features of correlation coefficients measured at a receiver and by an eavesdropper, we use the Monte Carlo experiment on 1000 trials.

The Monte Carlo simulation of bit identification process at a receiver and of the first step of identification for a would-be eavesdropper, which has taken an *S*-length part of the message as a replica, is shown in Fig. 1 for $S = 25$ from 50 samples of a bit-bearing packet. Fig. 1 a) shows the correlation coefficient of the replica and a bit-bearing signal measured at a receiver (upper set) and the background (lower set), which consists of the measurements of correlation coefficients of this replica with other signals. We can see here a very good separation of bit-identifying correlation coefficients (upper values) and the background coefficients (lower values). Fig. 1 b) shows the correlation coefficients measured by a bit-finder, which takes supposedly such part of the message which contains all samples of bit-bearing signal and noise samples. We can see here that in this case the measured correlation coefficient (solid black curve) is buried in the background noise.

Fig. 1 c) shows the most important case when a bit-finder takes smaller part of the message, let it be of *S*-length. With this part in hand, he scans the rest of the message with the goal of finding bits. Obviously, he wants to use the least possible length of such a part to find bits because the number of trials is proportional to C_{2S}^{S}. However, as we can see in Fig. 1 c), one cannot take sufficiently small *S*-length and, at the same time, identify bits with fidelity. Figure demonstrates why in the case of $S = 25$ such identification is impossible. Really, the correlation coefficient measured by a bit-finder (solid black curve) is buried partially in noisy background. These heuristic results can be justified on formal mathematical level. Accounting for normal distribution of correlation coefficients, it is easy to calculate the Type I error of identification (α) for both a receiver and a bit-finder. To this end, the mean correlation coefficients and theirs standard deviations are calculated by averaging all magnitudes exhibited in Fig. 1. With known statistical features in hand, we calculate the threshold magnitudes for receiver and bit-finder as the mean of corresponding bit-bearing and background correlation coefficients ($k^{(0,1)}$ for a receiver, and $\tilde{k}^{(0,1)}$ for a bit-finder) as one of possible methods.

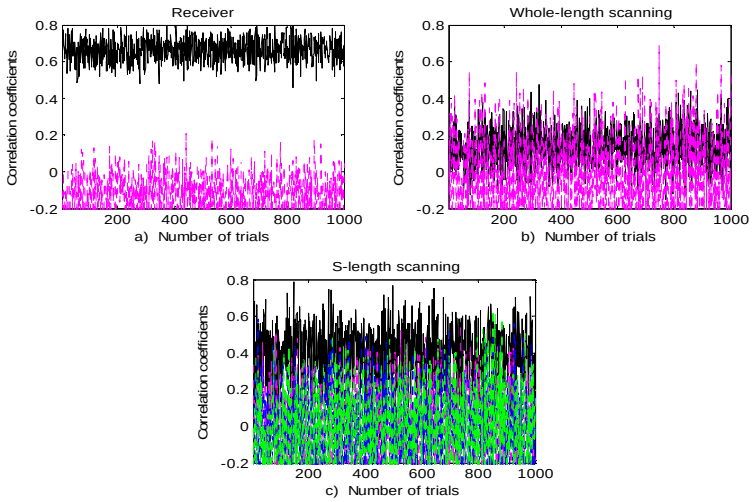

Fig. 1. Monte Carlo experiment on 1000 trials for $N = 50$, $SNR = 0.8$, $SNRN = 0.2$, and $S = 25$. a) The correlation coefficient of the replica and a bit-bearing signal measured at a receiver (upper set) and the background (lower set), which consists of the measurements of correlation coefficients of this replica with other signals. b) The correlation coefficient of a large part of the message taken by a bit-finder as a replica and other parts of the message. c) The correlation coefficient of an *S*-length part of the message taken by a bit-finder as a replica and other *S*-length parts of the message.

Then, an identification error at a receiver: $\alpha^{(0,1)} = 1 - P(r > k^{(0,1)})$, and for a bit-finder: $\tilde{\alpha}^{(0,1)} = 1 - P(r > \tilde{k}^{(0,1)})$. For considered case of $N = 50$, $SNR = 0.8$, and $SNRN = 0.2$, the identification error of bits identification at a receiver, $\alpha^{(0,1)} \sim 2.5 \cdot 10^{-12}$. As to a bit-finder, if he takes accidentally the whole 2*N*-length packet as a base signal for scanning, the identification will be not possible at all (see Fig. 1 b)). If a bit-finder takes sufficiently small 2*S*-length signal with $S = 25$ as a base one, he has to try $10^7 C_{50}^{25} \approx 1.2641 \cdot 10^{21}$ times for finding bits. However, all these trials are of little use because the identification error in each trial, $\tilde{\alpha}^{(0,1)} \sim 0.1$. Hence, while finding the needed distribution of bit-bearing samples, a bit-finder might be in error approximately $1.2641 \cdot 10^{20}$. This example shows why a cryptanalyst cannot find bits with fidelity.

CONCLUSION

Stochastic principle in comparing of electromagnetic signals or series of numbers in floating point format has been applied in this paper to the safe communication problem, which is tied with the safe computer problem. Both these problems can be solved with the aid of the new method of information distribution. For this, a text may be transformed into binary code and then each bit transformed into series of random samples. Such series may be electrical signals or numbers in floating point format. It is also possible the other scenario: letters of a text are transformed directly into the series of random samples. Both of the two are acceptable in different practical cases.

As has been shown in the work, there is no opportunity to decamouflage bits or letters with the aid of cryptanalysis of intercepted message. More to the point, the communication line may be so designed that no one intruder can send any bit-based information into the computer area. These merits are possible only at the expense of significant diminishing of information exchange rate.

REFERENCES

[1] Bruce Schneier. Applied Cryptography. 2nd edition. New York: John Wiley & Sons, 1995.

[2] Bruce Schneier. Secrets and Lies: Digital Security in a Network World. New York: John Wiley & Sons, 2000.

[3] Simon Singh. The Code Book. New York: Doubleday, 1999.

[4] Ham D., Li X., Denenberg S. A., Lee T. H., and Ricketts D. S. Ordered and Chaotic Electrical Solitons: Communication Perspectives // IEEE Commun. Mag. Dec. 2006. Vol. 44, no. 12, pp.126-34.

[5] Pecora L. M. and Carroll T. L. Synchronization in Chaotic Systems // Physical Rev. Lett. Feb.1990.Vol. 64, no. 8, pp. 821-25.

[6] Cuomo K. M. and Oppenheim A. V. Synchronization of Lorenz-based Chaotic Circuits with Application to Communication // IEEE Trans. Circuits and Sys. Oct. 1993. Vol. 40, no. 10, pp. 626-33.

[7] Roy R. and Thornburg K. S. Jr. Experimental Synchronization of Chaotic Lasers // Physical Rev. Lett. Mar. 1994, vol. 72, no. 13, pp. 2009-15.