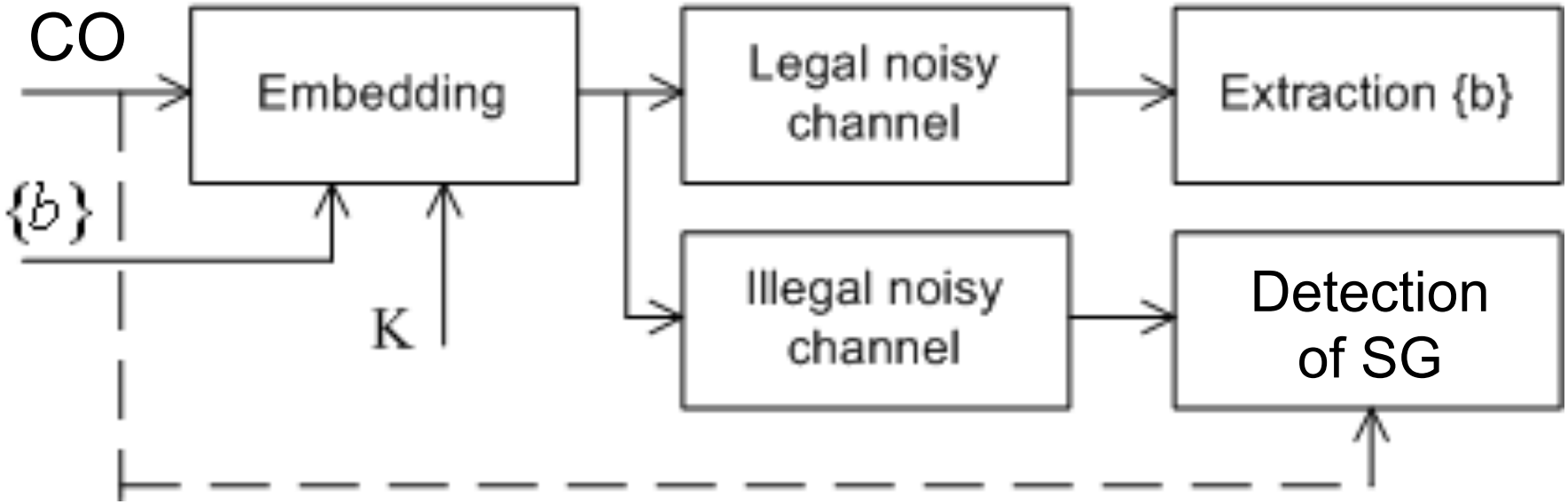
Lecture 6. SG based on noisy channels [42].



The peculiarities of the model:

- it is allowed to use by attacker only noisy version of SG $C_w(n)$,
- CO can be known exactly for attacker.

Practical applications:

- transmission of SG over satellite, mobile or optical fiber channels,
- interception SG by attacker over side channels,
- simulation of noisy channels.
- slightly noised musical files placed on some sites in the Internet.

The main idea to design SG based on noisy channels:

"Camouflage" the embedded message by noise of the channel.

Attacker's problem in this setting :
To distinguish if there exists channel noise only either both channel noise and SG.

Simplification in design of SG based on noisy channel in comparison with conventional case:
Description of distribution for channel noise is simpler than description of CO distribution .
We consider two types of the channels*:*

1. Binary symmetric channel without memory (BSC).
2. Continuous channel with Gaussian white additive noise.

Condition. We will assume that legal and illegal channel have the same distributions of noise.

1. SG based on BSC.

Embedding:

$$C_w(n) = C(n) \oplus \Theta_b(n)\pi(n), n = 1,2..N \quad (1)$$

$$C(n) \in \{0,1\}; \pi(n) \in \{0,1\}, i.i.d., P(\pi(n) = 1) = P_w, P(\pi(n) = 0) = 1 - P_w, \Theta_b(n) = b,$$

$$n = 1,2..N$$

"$\oplus$" - modulo two addition .

After a passing of $C_w(n)$ over BSC we get:

$$C'_w(n) = C_w(n) \oplus \varepsilon(n), n = 1,2..N \quad (2)$$

$$\varepsilon(n) \in \{0,1\}; i.i.d., P(\varepsilon(n) = 1) = P_0, P(\varepsilon(n) = 0) = 1 - P_0, P_0 < 1/2.$$

Attack in order to detect SG:

1. $\tilde{C}_w(n) = C'_w(n) \oplus C(n)$, ($C(n)$ – is known exactly by attacker) (3)
2. Two hypothesis testing:

$H_0 : \tilde{C}_w(n), n = 1,2..N$ - Bernoulli sequence with the parameter P_0,

$H_1 : \tilde{C}_w(n), n = 1,2..N$ - Bernoulli sequence with the parameter

$$P_1 = P_0(1 - P_w) + P_w(1 - P_0).$$

Evident testing method :

H_0, if $t < \tau$,

H_1, if $t \geq \tau$, (4)

where $t = H_w(\tilde{C}_w(n), n = 1,2..N)$, $H_w(\boldsymbol{X})$ – the Hamming weight of $\boldsymbol{X}$.

τ – some threshold .

Performance evaluation of SG detecting is determined by probabilities P_{fa} and P_m:

$$P_m = \sum_{i=1}^{\tau-1} \binom{N}{i} P_1^i (1-P_1)^{N-i}, P_{fa} = \sum_{i=\tau}^{N} \binom{N}{i} P_0^i (1-P_0)^{N-i}, \quad (5)$$

where $\binom{N}{i} = \frac{N!}{i!(N-i)!}$.

But the exact calculation by (5) is hard for $N > 10^3$. Thus we will use the criterion of relative entropy (see Lecture 5).

$$D = D\left(P_{H_0} \,\|\, P_{H_1}\right) = \sum_{x \in X} P_{H_0}(x) \log \frac{P_{H_0}(x)}{P_{H_1}(x)},$$

For BSC model where $X = \{0,1\}$, we get easily

$$D = N\left(P_0 \log \frac{P_0}{P_1} + (1 - P_0) \log \frac{1 - P_0}{1 - P_1} \right). \tag{6}$$

Then the bound for P_{fa} and P_m (see Lecture 5) is :

$$P_{fa} \log \frac{P_{fa}}{1 - P_{fa}} + (1 - P_{fa}) \log \frac{1 - P_{fa}}{P_m} \le D, \tag{7}$$

where D – is calculated by (6).

Next we will find the probability of error P_e under extraction of the embedded bit by informed legal decoder .

Decoding scheme:

$$\Lambda = \sum_{n=1}^{N_0} (C'_w(n) \oplus C(n)) \pi(n) \Rightarrow \tilde{b} = \begin{cases} 1, \Lambda \ge \lambda \\ 0, \Lambda < \lambda \end{cases}, \tag{8}$$

where λ – is some threshold, N_0 – is the number of samples which were used to embed one information bit.

Substituting (1) and (2) into (8), we get

$\Lambda = \Lambda_b$ for $b = 0$ and $b = 1$, where

$$\Lambda_0 = \sum_{n=1}^{N_0} \varepsilon(n)\pi(n), \Lambda_1 = \sum_{n=1}^{N_0} (\varepsilon(n) \oplus \pi(n))\pi(n). \quad (9)$$

Then

$$P(0/1) = \sum_{\delta=1}^{N_0} P(0/1,s)P(H_w(\pi) = s), \quad (10)$$

where $P(0/1,s) = P(\Lambda_1 < \lambda / s) = \sum_{j=s-\lambda}^{s} \binom{s}{j} P_0^j (1-P_0)^{s-j}$ (11)

$H_w(\boldsymbol{\pi})$ – is the Hamming weight of the sequence $\boldsymbol{\pi}(n)$, $n=1,2\ldots N$.

$$P(H_w(\pi) = s) = \binom{N_0}{s} P_w^s (1-P_w)^{N_0-s}. \quad (12)$$

Substituting (12) and (11) into (10) we get

$$P(0/1) = \sum_{s=1}^{N_0} \binom{N_0}{s} P_w^s (1-P_w)^{N_0-s} \sum_{j=s-\lambda}^{s} \binom{s}{j} P_0^j (1-P_0)^{s-j}. \quad (13)$$

In a similar manner we can obtain that

$$P(1/0) = \sum_{s=1}^{N_0} \binom{N_0}{s} P_w^s (1-P_w)^{N_0-s} \sum_{j=\lambda}^{s} \binom{s}{j} P_0^j (1-P_0)^{s-j}. \quad (14)$$

Because the sequence $\pi(n)$ is known at the decoder it is possible to get $P(0/1) = P(1/0)$, selecting $\lambda = S/2$, that gives

$$P_e = P(0/1) = P(1/0) = \sum_{s=0}^{N_0} \binom{N_0}{s} P_w^s (1-P_w)^{N_0-s} \sum_{j=s/2}^{s} \binom{s}{j} P_0^j (1-P_0)^{s-j}. \quad (15)$$

We apply Chernoff bound to inner sum in (15) in order to simplify it :

$$\sum_{j=s/2}^{s} \binom{s}{j} P_0^j (1-P_0)^{s-j} \le \left[4P_0(1-P_0)\right]^{s/2}, \quad (16)$$

and substituting (16) into (15) we get

$$P_e \le \sum_{s=0}^{N_0} \binom{N_0}{s} P_w^s (1-P_w)^{N_0-s} \left[4P_0(1-P_0)\right]^{s/2}. \quad (17)$$

Finally using Binomial Newton formula in (17) we have

$$P_e \le \left[(2\sqrt{P_0(1-P_0)} - 1)P_w + 1\right]^{N_0}, \quad \text{because} \quad (18)$$

$$\left((a+b)^{N_0} = \sum_{K=0}^{N_0} \binom{N_0}{K} a^{N_0-K} b^K \right).$$

Optimization problem in design of SG based on noisy channels :
Given D (security), P_0 (state of BSC), P_e (the desired reliability)
it is necessary to find the parameters P_w and N, that provide maximum of the number secure embedded bits $m = N/N_0$.

$$P_e \le \left[(2\sqrt{P_0(1-P_0)} - 1)P_w + 1 \right]^{N/m},$$

$$D = N\left(P_0 \log\frac{P_0}{P_1} + (1-P_0)\log\frac{1-P_0}{1-P_1} \right), \qquad P_1 = P_0(1-P_w) + P_w(1-P_0). \tag{19}$$

Example. *D=0.1, P_0=0.01, P_e=0.001*. Then *m=263* for *N=13739100, $P_w = 6.27 \cdot 10^{-7}$*.
If we take the restriction *$N \le 10^6$*, then *m=19* for *$P_w = 8.623 \cdot 10^{-6}$*.

2. SG based on Gaussian channels.

Embedding:

$$C_w(n) = C(n) + (-1)^b \sigma_w \pi(n), n = 1,2..N, \quad (21)$$

where $\pi(n) - i.i.d. \in N(0,1), \sigma_w$ – amplitude coefficient.

After a passing of SG $C_w(n)$ over Gaussian channel we get:

$$C'_w(n) = C_w(n) + \varepsilon(n), n = 1,2..N, \quad (22)$$

where $\varepsilon(n) - i.i.d. \in N(0,\sigma_\varepsilon^2)$.

Attack to detect SG consists in two steps:

1. $\tilde{C}_w(n) = C'_w(n) - C(n)$, ($C(n)$ – is known exactly for attacker).

2. Two hypothesis testing:

$$\begin{aligned} H_0 &: \tilde{C}_w(n), n = 1,2..N, i.i.d., \in N(0,\sigma_\varepsilon^2), \\ H_1 &: \tilde{C}_w(n), n = 1,2..N, i.i.d., \in N(0,\sigma_\varepsilon^2 + \sigma_w^2). \end{aligned} \quad (23)$$

There exists well known testing methods for this model but it is more convenient to apply relative entropy technique for two Gaussian distributions .

$$D = D\left(P_{H_0} \,\|\, P_{H_1}\right) = N\int_{-\infty}^{+\infty} P_{H_0}(x)\log\frac{P_{H_0}(x)}{P_{H_1}(x)}dx, \tag{24}$$

where $P_{H_0}(x) \in N(0,\sigma_\varepsilon^2), P_{H_1}(x) \in N(0,\sigma_\varepsilon^2 + \sigma_w^2).$

After a calculation of the integral in (24) and simple transforms we get :

$$D = 0,72N\left[\ln(1 + 1/\eta_w) - (1+\eta_w)^{-1}\right], \tag{25}$$

where $\eta_w = \sigma_\varepsilon^2 / \sigma_w^2$ (noise-to-watermark ratio (NWR)).

For large NWR, that provides high security of SGS, we have from (25)

$$D \approx 0,36\frac{N}{\eta_w^2}. \tag{26}$$

Express η_w, from (26) as a function of *N* and *D*:

$$\eta_w \approx 0.6\sqrt{N/D}. \tag{27}$$

Next we find the probability P_e under extraction of the embedded bit by legal informed decoder.

Decoding scheme.

$$\Lambda = \sum_{n=1}^{N_0} (C'_w(n) - C(n))\pi(n) \Rightarrow \tilde{b} = \begin{cases} 1, \Lambda \geq 0 \\ 0, \Lambda < 0 \end{cases}. \quad (28)$$

Substituting (21) and (22) into (28) we get

$$\Lambda = \sum_{n=1}^{N_0} (\varepsilon(n) + (-1)^b \sigma_w \pi(n))\pi(n). \quad (29)$$

The probability of error $P_e = P(0/1) = P(1/0)$ can be easily found by (29) taking into account that Λ – is Gaussian random variable owing Central Limit Theorem:

$$P_e = Q\left(\frac{|E\{\Lambda\}|}{\sqrt{Var\{\Lambda\}}} \right), \quad (30)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt.$

After simple calculation of *E{Λ}* and *Var{Λ}* we get

$$P_e = Q\left(\frac{\sqrt{N_0}\sigma_w}{\sqrt{\sigma_\varepsilon^2 + \sigma_w^2}}\right) \approx Q\left(\frac{\sqrt{N_0}\sigma_w}{\sigma_\varepsilon}\right) = Q\left(\sqrt{N_0/\eta_w}\right). \quad (31)$$

Substituting (27) into the last equality we obtain

$$P_e = Q\left(1.29\sqrt{(ND)^{1/2}/m}\right). \quad (32)$$

Remark. There is a simple bound for the function *Q(x)* [35]:

$$Q(x) \le \exp(-x^2/2). \quad (33)$$

Substituting (33) into (32) we get

$$P_e \le \exp\left(-0.83(ND)^{1/2}/m\right). \quad (34)$$

Example. Given *D=0.1, N=1000* we get $P_e \le 3\cdot 10^{-4}$.

Important remark. In Gaussian noisy channel it is possible to embed any number of bits *m* given any security level *D* and any reliability P_e.(In contrast to BSC case!)

General conclusions for SG based on noisy channels.

1. This is unique case where it is possible to provide a secrecy of SG in attack with known exactly CO.

2. A choice of channel model (BSC or Gaussian) depends on the condition in which place, of telecommunication line is allowed to embed and extract additional information .

3. Design of SG does not require the knowledge of CO statistics.

4. In the case of BSC the number of secure and reliably embedded bits is limited , whereas in the Gaussian case it is unlimited (at the coast of increasing of N) but the embedding rate approaches to zero as $N \to \infty$.

5. The use of error correcting codes allows to increase slightly the embedding rate but it is kept as before approaching to zero as N approaches to infinity on the contrary to conventional communication systems where due to Shannon's theorem the date rate can be fixed if it is less than channel capacity.
(See Lecture 15 "Capacity of SG and WM systems").

6. The more is the number of embedded bits m in the Gaussian case the less is amplitude of embedding that results in turn in problems for practical implementation.

Spread-time stegosystems (SG-ST) [42].

$$C_w(n) = C(n) + (-1)^b \sigma_w \pi(n) \text{ with the probability } P_0$$
$$C_w(n) = C(n) \text{ with the probability } 1 - P_0 \qquad (35)$$

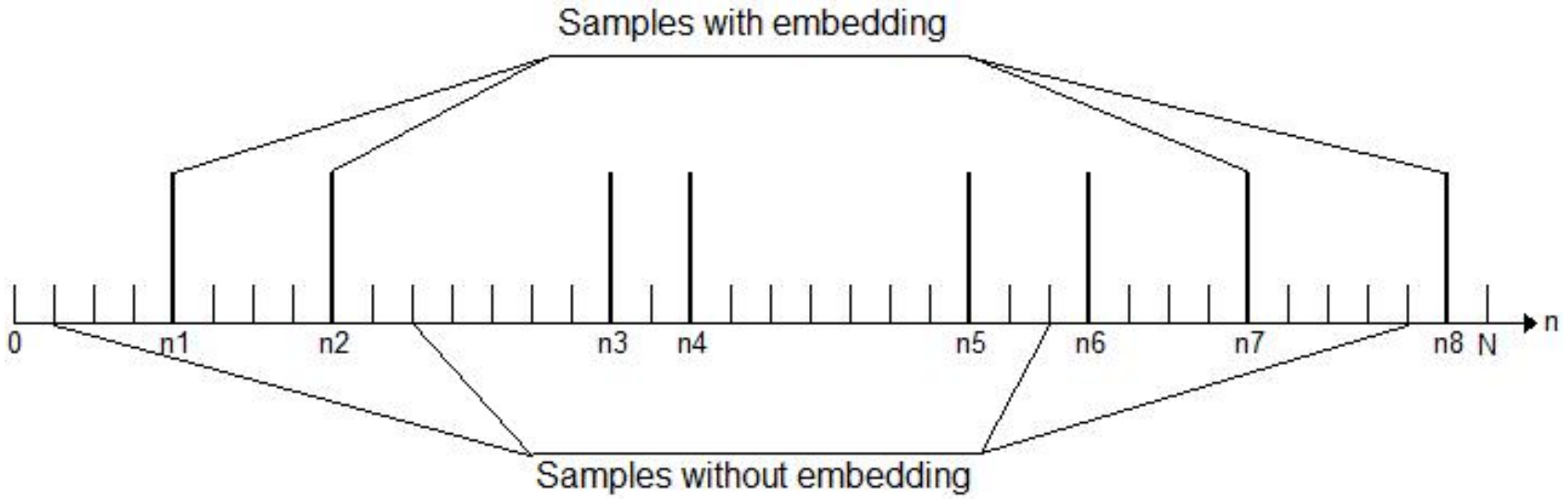


Fig 1. SG-ST system with embedding into pseudorandom samples. (Here $N_s = 8,\ N = 38,\ P_0 \approx 8/38$)

In order to take a decision about presence or absence of SG under the condition of known $C(n)$ he (or she) has to perform a testing of two hypothesis:

$$H_0 : \left\{ \delta(n) \in N(0, \sigma_\varepsilon^2), \delta(n) \in i.i.d. \right\}$$

$$H_1 : \begin{cases} \delta(n) \in N(0, \sigma_s^2), \delta(n) \in i.i.d. \text{ with the probability } P_0 \\ \delta(n) \in N(0, \sigma_\varepsilon^2), \delta(n) \in i.i.d. \text{ with the probability } 1 - P_0 \end{cases} \qquad (36)$$

$$\delta(n) = C'_w(n) - C(n),\ \sigma_s^2 = \sigma_\varepsilon^2 + \sigma_w^2$$

Optimal hypothesis testing based on maximum likelihood ratio is

$$\Lambda\left(\Lambda_1/\Lambda_0\right) \geq \lambda_0 \Rightarrow H_1,\ \Lambda\left(\Lambda_1/\Lambda_0\right) < \lambda_0 \Rightarrow H_0,$$

$$\text{where } \Lambda\left(\Lambda_1/\Lambda_0\right) = \frac{P(\boldsymbol{\delta}/H_1)}{P(\boldsymbol{\delta}/H_0)} \tag{37}$$

Suboptimal testing:

$$\left.\begin{array}{l} \tilde{\Lambda} \geq \tilde{\lambda}_0 \Rightarrow H_1,\ \tilde{\Lambda} < \tilde{\lambda}_0 \Rightarrow H_0, \\ \text{where } \tilde{\Lambda} = \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \delta^2(n). \end{array}\right\} \text{This method is intuitively reasonable.} \tag{38}$$

The efficiency of SG-ST can be estimated (as usually) by two probabilities:

P_m – the probability of SG - ST missing,

P_{fa} – the probability of SG - ST false detection (false alarm).

It was proved in [42] that if we let (for simplicity) that $P_m = P_{fa} = P,$ then the following lower bound for P holds true:

$$P \geq Q\left(\sqrt{\frac{N}{2}}\frac{P_0}{2\eta_w}\right) \text{ or } P \geq Q\left(\frac{N_s}{2\sqrt{2N}\eta_w}\right), \tag{39}$$

where $\eta_w = \frac{\sigma_c^2}{\sigma_w^2}$, $N_s = P_0 \cdot N$ (the total number of samples with embedding)

If $N_s \sim \sqrt{N}$ as $N \to \infty$, then $P \to 1/2$ and hence SG - STapproaches to undetectable SG.

The probability of bit error after its extraction by legal user (the relation (31))works.

Example. $N = 10^7, \eta_w = 20, N_0 = 210, P_0 = 0.0045 (N_s = 45260).$

Then 215 bits can be embedded with the bit error probability at most 10^{-3} and $P \geq 0.4$

The use of error correcting codes in SG-ST

Then an embedding procedure has to be performed as follows:

$$C_w(n_j) = C(n_j) + (-1)^{b_{ij}} \sigma_w \pi(n_j) \text{ with the probability } P_0$$
$$C_w(n_j) = C(n_j) \text{ with the probability } 1 - P_0 \qquad (40)$$

where b_{ij} denotes the j - th bit the i - th codeword of the lenght $N_0 = N/l$,

l – is some integer value.

Decoder takes a decision about the embedding of the i -th codeword by making

$$i = \arg\max_{1 \le i \le 2^k} \sum_{n=1}^{N_0} (C'_w(n) - C(n))(-1)^{b_{ij}} \pi(n) \qquad (41)$$

The probability of block error P_{be} based on well known union bound [43] can be expressed as

$$P_{be} \le (2^k - 1) Q\left(\sqrt{\frac{d}{\eta_w}}\right) \le \exp\left(-\frac{d}{2\eta_w} + RN_0 \ln 2\right),$$

where $R = k/N$

Example. $N = 10^7, P \ge 0.4, \eta_w = 20, N_0 = 210, P_{be} \le 10^{-3}.$

Then using simplex code (1023,10,512) it is possible to embed 442 bits.

Optimal detecting of SG-ST

Following to eq.(37) and to notations (36) we get the logarithmic normalized likelihood ratio as

$$\Lambda_L(\Lambda_1 \mid \Lambda_0) = \frac{1}{N}\sum_{n=1}^{N} \log\left[P_0 \exp\left(\frac{1}{2\eta_W^2 \sigma_\varepsilon^2}\, \delta(n)^2 \right) + (1 - P_0) \right] \qquad (42)$$

Since N is sufficiently large, we can apply the Central Limit Theorem to the sum in (42).Then we get for such choice of the threshold λ, which provides $P_m = P_{fa} = P$ the following upper bound

$$P \geq Q\left(\frac{\tilde{\mu}_1 - \tilde{\mu}_0}{2\tilde{\sigma}_0} \right) \qquad (43)$$

where, for $j = 0,1$

$$\tilde{\mu}_j = E\left[\left(\log\left(P_0 \exp\left(\frac{\delta(n)^2}{2\eta_W \sigma_\varepsilon^2} \right) + \left(1 - P_0\right) \right) \right)_{n=1}^{N} \middle| H_j \right]$$

$$\tilde{\sigma}_0 = \frac{1}{N} \mathrm{Var}\left(\left(\log\left(P_0 \exp \frac{\delta(n)^2}{2\eta_W \sigma_\varepsilon^2} \right) + \left(1 - P_0\right) \right)_{n=1}^{N} \middle| H_0 \right) =$$

$$= \frac{1}{N}\left(E\left[\left(\left(\log\left(P_0 \exp \frac{\delta(n)^2}{2\eta_W \sigma_\varepsilon^2} \right) + \left(1 - P_0\right) \right)_{n=1}^{N} \middle| H_j \right) \right] - \tilde{\mu}_0^2 \right)$$

where the random values $\delta(n)$ have the probability distributions given by (36).

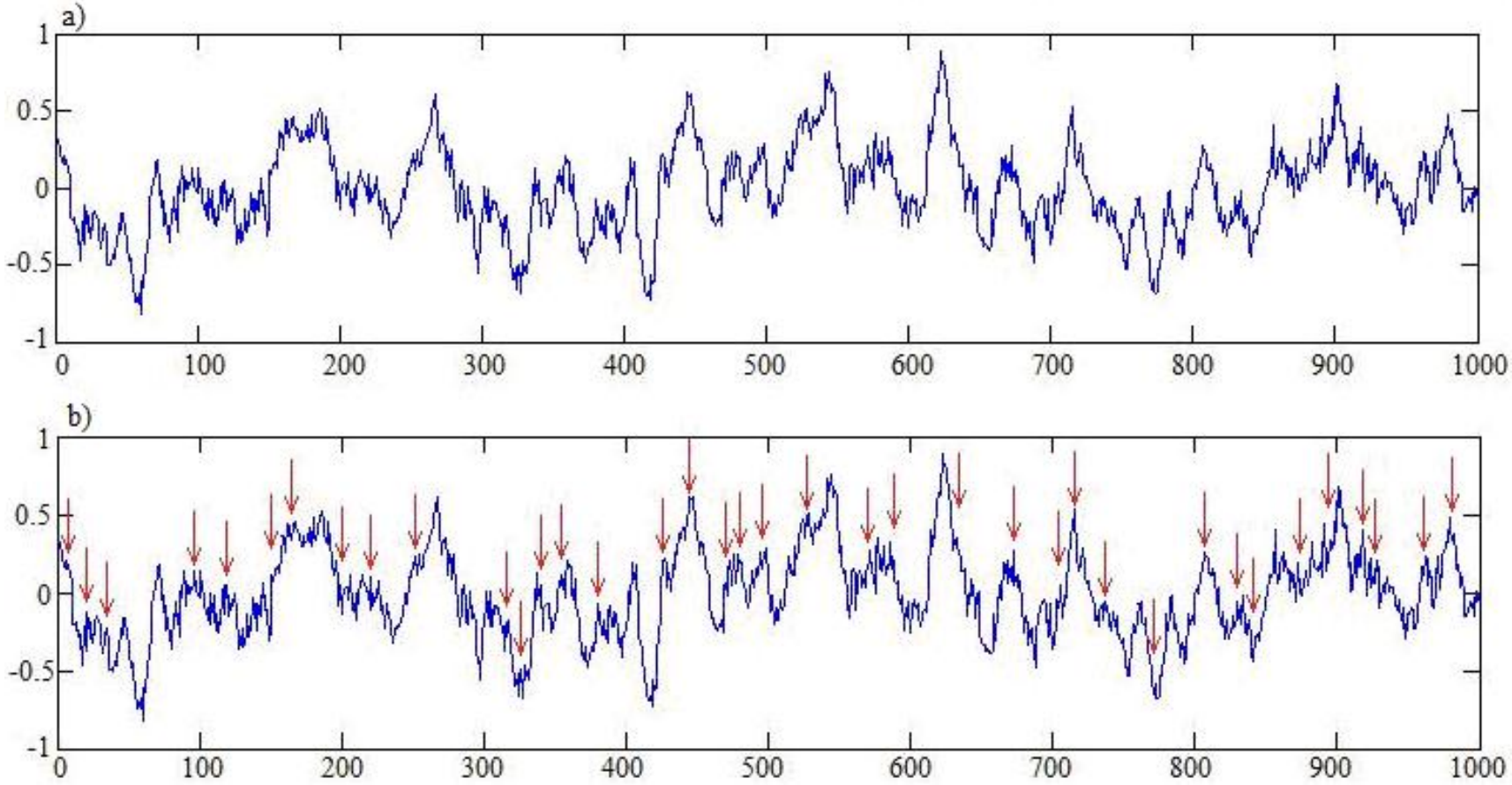
Since it is very hard to find analytically the values $\tilde{\mu}_0, \tilde{\mu}_1$ and $\tilde{\sigma}_0$, we estimate them just by the simulation.

N	σ_ε^2	η_w	$P_0 = N_S/N$	$\tilde{m}_0$	$\tilde{m}_1$	$\tilde{\sigma}_0$	$P = Q\left(\frac{\tilde{m}_1 - \tilde{m}_0}{2\tilde{\sigma}_0}\right)$
10^4	1	20	0.1431	0.00161414	0.00162667	0.00240398	0.401753
		50	0.3578	0.00157674	0.00158801	0.00229017	0.401759
		100	0.7156	0.00156462	0.00157585	0.00225445	0.401737
	5	20	0.1431	0.00161414	0.00162590	0.00240398	0.401754
		50	0.3578	0.00157675	0.00158821	0.00229017	0.401745
		100	0.7156	0.00156462	0.00157583	0.00225445	0.401726
10^5	1	20	0.04526	0.000512618	0.000513737	0.000767001	0.401741
		50	0.1131	0.000500332	0.000501449	0.000729712	0.401809
		100	0.2263	0.000496672	0.000497830	0.000718483	0.401737
	5	20	0.04526	0.000512618	0.000513854	0.000767001	0.401772
		50	0.1131	0.000499062	0.000500277	0.000727769	0.401806
		100	0.2263	0.000496672	0.000497795	0.000718483	0.401745

N	σ_ε^2	η_w	$P_0 = N_S/N$	$\tilde{m}_0$	$\tilde{m}_1$	$\tilde{\sigma}_0$	$P = Q\left(\frac{\tilde{m}_1 - \tilde{m}_0}{2\tilde{\sigma}_0}\right)$
10^6	1	20	0.01431	0.000162288	0.000162393	0.000243341	0.401835
		50	0.03578	0.000158479	0.000158585	0.000231440	0.401862
		100	0.07156	0.000157686	0.000157808	0.000228397	0.401548
	5	20	0.01431	0.000162288	0.000162435	0.000243187	0.401752
		50	0.03578	0.000158479	0.000158598	0.00023144	0.401711
		100	0.07156	0.000157686	0.000157797	0.000228397	0.401461
10^7	1	20	0.004526	$5.13502\cdot10^{-5}$	$5.13615\cdot10^{-5}$	$7.69844\cdot10^{-5}$	0.401964
		50	0.01131	$5.01145\cdot10^{-5}$	$5.01246\cdot10^{-5}$	$7.32173\cdot10^{-5}$	0.401900
		100	0.02263	$4.97464\cdot10^{-5}$	$4.97587\cdot10^{-5}$	$7.20836\cdot10^{-5}$	0.401777
	5	20	0.004526	$5.13502\cdot10^{-5}$	$5.13626\cdot10^{-5}$	$7.69844\cdot10^{-5}$	0.401812
		50	0.01131	$5.01145\cdot10^{-5}$	$5.01245\cdot10^{-5}$	$7.32173\cdot10^{-5}$	0.401969
		100	0.02263	$4.97464\cdot10^{-5}$	$4.97569\cdot10^{-5}$	$7.20836\cdot10^{-5}$	0.401686

It can be seen that the use of the optimal decision rule does not break undetectability of SG-ST, hence it can be declared as a secure SG system indeed.

The waveforms of audio signal after its passing over noisy channel with signal-to-noise ratio 10 dB (a), and the wave form of the same signal after embedding by SG-ST algorithm with NWR=20dB (b).
(The arrows show the samples with embedding)



We can see that neither visually nor by ears the presence of some embedding cannot be detected.
The fact of undetectability of SG-ST by optimal statistical methods has been proved before.