Gouxi CHEN[1,2], Meng ZHANG[1], Junjie CHEN[2], Donglai FU[1], Yuliang WU[1]

North University of China(1), Taiyuan University of Technology (2)

Capacity and security for imperfect batch steganography

Abstract. *Batch steganography aims to conquer the limitations of capacity and security of data hiding with a single cover, and it is to spread secret information among a large number of covers, having improved basic algorithm and strategies of stegosystem. The steganographic features of batch covers and single cover were compared to propose a theoretic model for batch steganography. According to Kullback-Leibler divergence of batch steganography, the relationship of imperfect steganographic capacity with a formalized number of batch covers and embedding strategies was put forward. The maximum payload of an imperfect batch steganographiy (IBS) was determined by the number of batch covers and the base cover set and proved by formulations, and compared with capacity of noisy channels. The results show that the model under constraints of IBS is feasible for batch steganography.The research results are helpful for stegosystem design and steganalysis.*

Streszczenie. *Artykuł analizuje możliwości steganografii – metody szyfrowania nie tylko treści, ale także ukrywania w ogóle faktu obecności przesyłu wiadomości. (**Możliwości i bezpieczeństwo niedoskonałej steganografii pakietowej**)*

Keywords: Steganography, Steganographic capacity, security, Steganalysis.
Słowa kluczowe: steganografia, bezpieczeństwo danych

.

Introduction

The classic covert communication model is the "Prisoners' Problem" proposed by Simmons [1]. In the model, Alice and Bob communicate without being detected by Warden Willie. Its focus is the security of the stegosystem and steganographic capacity [2,3]. Previous studies concentrated on efficient embedding and limitations of data into a single cover. Steganalysis can reduce the embedding capacity of steganography with a single cover, a method to increase steganographic capacity is batch steganography, which spreads secret information among a large set of covers. Ker [4] have proposed steganographic methods and segmented the basic theory of batch steganography. Fridrich [5,6] has proposed the methods of estimation capacity and a formal relationship [7] between maximum payloads and a number of covers on the security of stegosystems. Another research improved limitations, proposing a judged principle on the security of batch steganography, and establishing a theoretic foundation on hiding information [8].

These studies proposed theories for batch steganography but did not describe a basic theoretic model and essential condition of embedding strategies. They also did not present constraints on stegosystem security.

This paper investigates the basic theory of Kullback-Leibler (K-L) divergence [9], and proposes a theoretic model of an imperfect batch stegosystem and the constraints of security on a steganographic ε-secure[10] definition. Giving covers of uniform capacity and a quantitative steganalysis method satisfying a hypothesis test and certain conditions of assumptions, the relationship and security were proved by formulation. Constructing a batch steganography model of information hiding, and proved system security, improve the ability of defense steganalysis.

The structure of this paper is as follows. Section 1 explains the theoretic models of batch steganography. Section 2 describes the capacity and security analysis of batch stegosystems with imperfect condition. Section 3 proves the relative equality for K-L divergence under certain assumptions. Section 4 is the experiment and result analysis, and finally is the conclusion.

Model for batch steganography

Basic theoretic models of steganography are a noisy channel model[10] and the "Prisoners' Problem" Simmons has been proposed. However, they do not propose a theoretic model on batch steganography or its basic definitions. Shannon's Second Theorem coding theory stated the constraints of noisy channel capacity, ratio of translation, and channel coding. Depending on Nyquist theory and Shannon's theorem, Gaussian noisy channel states that the channel capacity of low signal-to-noise ratio (SNR) increases the noise of a system as it relates to channel capacity, bandwidth, and SNR. Ramkumar designed a steganographic channel model with noise based on this model. Simmons proposed a definition of a subliminal channel in narrow and wide senses, and further expressed a basic model of covert communication based on the subliminal channel. This scheme mainly uses fingerprints.

Previous research described steganography and steganalysis for single cover (digital media, e.g., image). This paper will present a theoretic model for batch steganography. In Fig.1, the model finish data has large hiding capacity, improves the security of the steganographic scheme, and increases the competence of fortification for corresponding steganalysis.

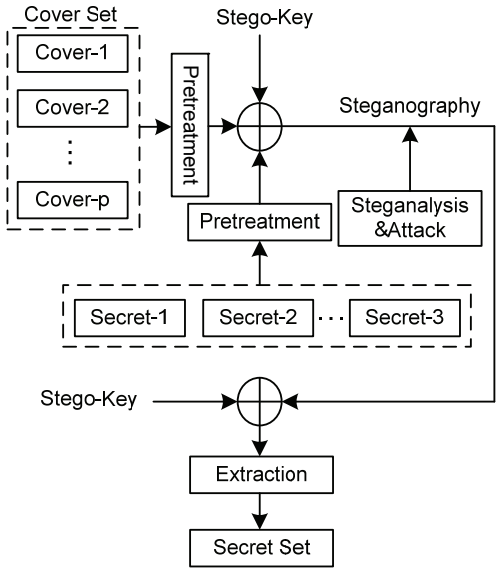

Fig.1. Batch Steganographic model

The model of batch steganography is composed of the following:

(1) Sorter of covers set
(2) Pretreatment of secret information
(3) Splitting secret information
(4) Embedding algorithm

(5) Extraction algorithm , and so on

Batch steganography generalizes the problem of hiding data M to N covers. The focus of this paper is embedding strategies, maximum payload, and the security of the stegosystem.

Basic assumption and defintions

DEFINITION 1. Suppose:

(1) C is covers set: $C=\{c_1,c_2,\ldots,c_a\}$and a is the number of covers. If $C_1,C_2,\ldots C_N$ is the sequence set, write $\{C_n\}$, n is a positive integer and $n\le N$ s.t. $C_i=\{c_k\,|0<k\le i\}$.

(2) S is a set of secret information blocks:

$S=\{s_1,s_2,\ldots,s_q\}$,for the size of S, write $|S|$ as: $|S|=\sum_1^q|s_j|$ .q is the number of secret information blocks.

(3) C_b is the steganographic covers set that must satisfy:

$C_b\subseteq\{C_n\},|C_b|>|S|$ $(0<b\le a)$

write $N=\overline{\overline{C_b}}$, which is the number of aiming steganographic covers.

(4) K is a set of all stego-keys $K=\{k_1,k_2,\ldots,k_v\}$, v is the number of stego-keys.

(5) C_S are covers with S and K.

(6) E_K is an embedding algorithm:$C\times S\times K\to C_S$.

(7) D_K is an extraction algorithm of secret information: $C_S\times K\to S$. $D_K(E_K(c_j,s_i,k))=s_i$ s.t. $\forall\, s_i\in S\mid 1\le i\le q,\ c_j\in C\mid 1\le j\le N$.

Thus, the stegosystem that satisfies the steganographic condition of (1)-(7) is defined as batch stegosystem, written as:

(1) $T_m=(C,S,K,C_S,E_K,D_K)$.

Then, the aforementioned definition such that: (1) The basis of sequence sets $\{C_n\}$ is monotonic increasing, and chosen strategies from covers set C to C_b is a complex algorithm. (2) S is a set of secret information, as well as a large amount of secret information blocks. (3) Stego-key set K includes the scheme of private and public keys.(4) Embedding algorithm E_K and extraction algorithm D_K converse mutually. (5) Batch stegosystem T_m is a function of six groups.

DEFINITION 2. In batch stegosystem T_m, hiding information of batch covers C_b must satisfy condition R_m, such that: (1) Intangibility for cover set, (2) Security of stego-key set, (3) Robustness for T_m.

Therefore, maximum payload M of batch steganography is defined as:

$|M|=\{m|m\in M,M|\,(R_m,\Gamma_m)\}$,

satisfying the following constraints:

(1) Cover set is $C_b=\{c_1,\ c_2,\ldots,c_p\}$, p is the number of covers.

$C_b\subseteq\{C_n\}$ and $|C_b|>|S|$ （$0<b\le a$）.

(2) C_b satisfies R_m of maximum payload M for the entire stegosystem.

(3) q is a secret blocks' number for the corresponding maximum payload M, $|M|=\sum_1^q|m_j|$.

(4) Δf is the changed rate of C_S based on visible features, and satisfying $\Delta f\le JND$. Just noticeable degree (JND) is a visible threshold.

Thus, the yielded relative model with regard to batch covers and corresponding maximum payload is:

(2) $|M|=\max(g(C,EK,DK,R1,R2,R3,R4))$

Therein: $|M|$ is the maximum payload of batch covers C. R_1 is the relative coefficient of corresponding covers. R_2 is the relative coefficient in different embedding algorithm. R_3 is the relative coefficient of secret information. R_4 also is the relative coefficient of stego-key. g is a mapping, and satisfies:

(3) $(C,E_K,D_K)\times(R_1,R_2,R_3)\to|M|$

Certain assumptions can prove[11] THEOREM 1 as follows from DEFINITION 1 and 2.

THEOREM 1.Satisfying under condition for batch stegosystem Γ_m:

(1) S is secret information set, C_b is covers set hiding information, satisfying: $C_b\subseteq\{C_n\}$ $(0<b\le N)$ and $|C_b|>|S|$.

(2) The relationship of $c_1,c_2\ldots c_i$ $(0\le i\le p)$ of C_a in accordance with the static Markov Chain.

(3) $\forall$ $i,j(0\le i,j\le N,i\ne j)$, $R(E_i,E_j)=0$, E_i and E_j denote embedding practices of secret information for c_i and c_j of covers, respectively.

(4) $|M|=\sum_1^q|m_j|$, $|M|$ is maximum payload of hiding aimed covers.

Therefore,

(1) If $|M|/\sqrt{N}\to\infty$ as $N\to\infty$ then for sufficiently large N, Γ_m is at risk.

(2) If $|M|/\sqrt{N}\to 0$ as $N\to 0$ then for sufficiently large N, Γ_m is safe.

Thus, the foregoing conclusion explains that the noisy channel theory does not apply estimation capacity and constraints of security. Relative to the steganography of a single cover, steganographic security is proportional to the dispersed degree of secret information. However, for batch steganography, under certain assumptions, its security is not linear with the dispersed degree of secret information, depending on the square root of number of covers. Namely, increasing the ratio of $\sqrt{N}$ is much faster than increasing the ratio of $|M|$.

DEFINITION 3. Supposing batch stegosystem T_m satisfies:

(1) The steganographer knows hiding methods and has infinite calculating resources.

(2) The quantity of covers set C and aimed hiding covers set C_b are suffici-ently large.

Then, this T_m is defined as perfect batch ste-gosystem (PBS).

If T_m does not satisfy foregoing constraints, T_m is defined as IBS.

DEFINITION 4. If T_m does not satisfy foregoing constraints, T_m is defined as IBS. Suppose: T_m is batch stegosystem, P_{C_b} is density function of C_b , C_b = $\{C_1,\ C_2,\ldots,C_m\}$, m is the number of covers, and P_S is the density function of C_a carrying secret information. If $D(P_{C_B}\|P_S)\le\varepsilon$, then Γ_m is called ε-secure. Moreover, when ε = 0, T_m is perfectly secure.

Imperfect batch seganography (IBS)

Capacity of IBS

Warden Willie's task is to determine the maximum payload for which risk of detection is acceptable by methods of pooled steganalysis for given N covers

$C_S=\{C_S^1,C_S^2,\cdots,C_S^N\}$ C_S^i $(0<i\le N)$

treated by T_m, Willie's aim is to detect steganography, that is, to perform the hypothesis test:

$H_0:C_S^i=0$,

$H_1:C_S^i>0$,otherwise

$P=\{X\le x\mid C_S^i>0\}$ is the density function of T_m, and the element P_I of $P=\{P_1,P_2,\ldots,P_N\}$ corresponds with cover c_i and C_S^i carrying secret information. Suppose the quantity of C_S^i

under H_0 is A , the quantity of C_S^i under H_1 is B . Thus, steganalysis based on H_0 and H_1 are two Bernoulli trials, steganalysis for T_m are also Bernoulli trials. Under DEFINITION 1, supposing: Covers set $C_b=\{c_1,c_2,\ldots,c_N\}$ (N is the number of covers), $|c_i|=|c_j|(0\le i, j\le N)$,given:

$$\left|C_b\right|\sum_1^N P_i=\left|M\right|,\ |C_b|=N\,|c_1|=N\,|c_i|$$

Under H_1, $P_i=|s_i|/\left|C_S^i\right|$.

If $\gamma=p_i(0<\gamma<1)$, the embedding probability is 1-γ under H_0.

THEOREM 2. If T_m satisfies ε or ε^2-secure and cert-ain assumptions, then the steganographic probability of an imperfect stegosystem is determined by the number of covers and the number of embedding covers.

Proof: Following precious assumptions of (1) and (2), given for T_m:

$$(4)\quad P=\{X\le x \mid C_S^i>0\}=\binom{N}{B}(\gamma)^B(1-\gamma)^A\ ,A+B=N$$

$$(5)\ P=\frac{N!}{(N-B)!B!}(\gamma)^B(1-\gamma)^A=\frac{N!}{(N-B)!B!}\left(\frac{|M|}{B|C|}\right)^B\left(1-\frac{|M|}{B|C|}\right)^{N-B}$$

Where Γ_m is not at risk, following ε or ε^2-secure of THEOREM 1. In the literature[4], given :

$|M|=\omega\sqrt{|N|}$ (ω is a constant)

$$(6)\quad P=\frac{N!}{(N-B)!B!}\left(\frac{\omega\sqrt{N}}{B|C|}\right)^B\left(1-\frac{\omega\sqrt{N}}{B|C|}\right)^{N-B}$$

Such that $P=f(N,B,C)$ is a function of three groups with regard to N,B,C.

Deduction:

When fixing the number of covers and the number of steganographic aimed covers, max- imum payload is determined for an imperfect stegosystem, satisfying Theorems 1 and 2.

The aforementioned deduction showed that the maximum payload of IBS determined by the number of embedding covers and the number of all covers, has nothing to do with the embedding algorithm. That is, embedding strategies and covers set determine the maximum payload of the system.

Security of IBS

Suppose batch steganographic strategies are simply a separation equivalent, then $p=|M|/(N|C_b|)$, $p_i=p(0<i\le N)$. Assuming: Q_1 and Q_2 are continuous random variables, and q_1 and q_2 are corresponding density functions. Then type-I error of steganalysis is α , and type-II error of steganalysis is β. Following the definition of K-L div-ergence and ε-secure of Cachin, given

$$(7)\quad D_{KL}(Q_1||Q_2)=\int q_1(x)\frac{q_1(x)}{q_2(x)}dx$$

$$(8)\quad D_{KL}(\alpha,\beta)=\alpha\log_2\frac{\alpha}{1-\beta}+\beta\log_2\frac{\beta}{1-\alpha}\ (\alpha+\beta<1).$$

The determined process does not increase K-L div-ergence of the two kinds of distribution[16]. If $g:Q\to T$ then $D_{KL}(T_1||T_2)\le D_{KL}(Q_1||Q_2)$.

Not only is the hypothesis test on H_0 and H_1 Bernoulli trials for α and 1-β, but also a uniform distribution of probability p is based on covers hiding the information.

Suppose:

$f_1=\inf(\alpha_1,\alpha_2,\ldots,\alpha_N)$ and $f_2=\sup(\beta_1,\beta_2,\ldots,\beta_N)$,

f_1 and f_2 are the corresponding distributed functions of F_1 and F_2 :

$$(9)\quad D_{KL}(F_1||F_2)=N\int f_1(x)\log_2\frac{f_1(x)}{f_2(x)}dx$$

$$=N\int f_1(x)\log_2\frac{f_1(x)}{f_1(x-p)}dx$$

$$=N\int f_1(x)(\log_2 f_1(x)-\log_2 f_1(x-p))dx$$

$$=N\int f_1(x)\left[p\frac{f_1'(x)}{f_1(x)}-\frac{p^2}{2}f_1(\xi)''\right]dx$$

$$=Np\int f_1'(x)dx-N\frac{p^2}{2}f_1''(\xi)\int f_1(x)dx$$

$$=-\frac{p^2N}{2}f_1''(\xi)=-\frac{|M|^2}{N|C|^2}f_1''(\xi)$$

Such that $D_{KL}(F_1||F_2)=-\frac{|M|^2}{N|C|^2}f_1''(\xi)\le C^1\frac{|M|^2}{N|C|^2}$.

Suppose $C^1\ge -\frac{f_1''(\xi)}{2}$,($x-p<\xi<x$). Following THEOREM 1,given:Where $\sqrt{|M|}/N\to 0$, $\lim_{N\to\infty} D_{KL}(F_1||F_2)\to 0$

Because $D_{KL}(F_1||F_2)\ge 0$ (if $F_1=F_2$, equality exists) and $D_{KL}(F_1||F_2)$ is a convex function. If $N\to 0$ and $\sqrt{|M|}/N\to 0$, Then $D_{KL}(F_1||F_2)=0$

Stegosystem Γ_m is an imperfect stegosystem because of assumptions satisfying the statistical Markov Chain. Thus, Γ_m satisfies the ε-secure definition of Cachin.

Experimental results

Stegosystem Γ_m satisfy the conclusions of Theorem 1 and 2 under certain assumptions. Three kinds of strate-gies were adopted in the experiment of Embedding hidden information:(1) Embedding hidden information in fixed-size;(2) Embedding the secret information is proportional to $\sqrt{N}$ (N is the total number of pixels of covers);(3) Embedding the secret information in proportion to the total number N of pixels of covers(Images library from the NRCS web site), noncompressed images, and make a simple split into smaller equal bocks.

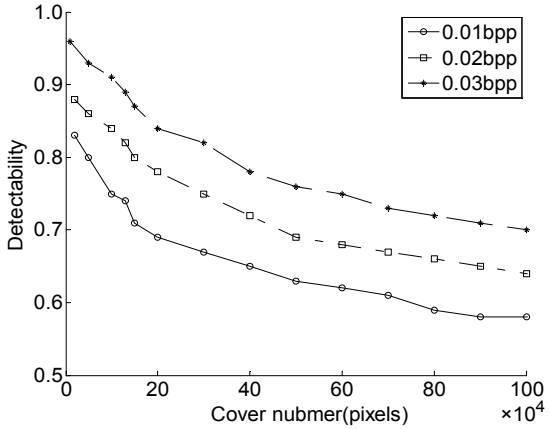

Fig.2. Relation of cover number and embedding capacity on fixed-size secure data

Fig2, Fig3 and Fig4 correspond to the experimental results of three embedding algorithm which are all used LSB method. In the experiment, maximum payload of batch steganography only strategy (2) exemplifies the case of regularity, the other consistent with the general analysis. Fig3 shows the curve illustrates the conclusion[8] must satisfy assumptions 1-4 in THEOREM 1.

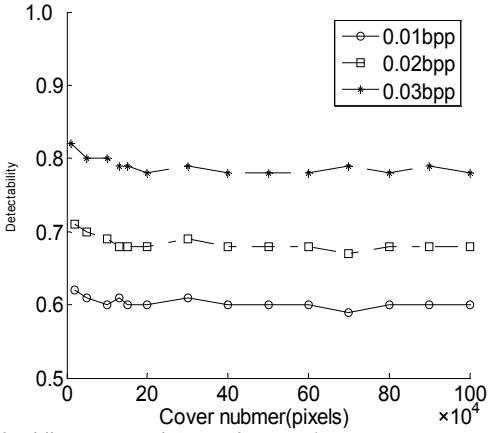

Fig.3. Embedding secure data on theorem1

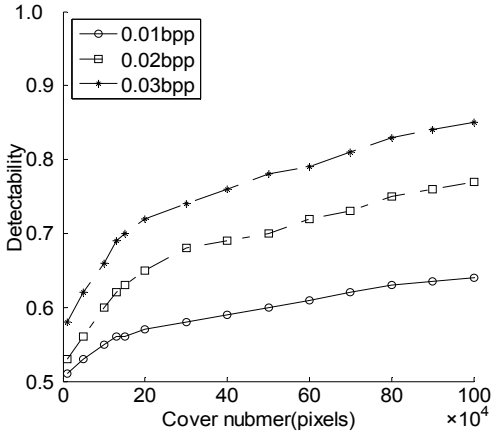

Fig.4.Embedding secure data proportional to cover number

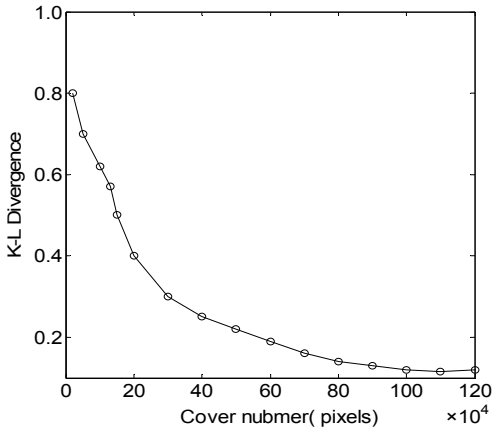

Fig.5. Security of imperfect batch stegosystem

Security of IBS in Fig5 can be verified. Satisfying certain assumptions, with the increase in the number of covers, when $N\to\infty$ and $\sqrt{|M|}/N\to 0$, D_{KL} gradually reduc-ed which reflects the security of steganographic systems analysis and forma proof of the experimental consistency, and also shows batch stegosystem security constraints exist legitimacy.

Conclusions

This paper proposed a batch steganographic theoretic model, and defined batch stegosystem and maximum payload for limitations of single steganography. It proved that batch steganographic capacity is different from noisy channel capacity theory. Under hypothesis test and certain assumptions, it obtained two kinds of relationships. One is batch steganographic strategies and capacity, and the other is K-L divergence. It proved that IBS is safe under certain constraints. Assumptions in the study were strict, thus, further research should study batch steganography under relaxed conditions.

Acknowledgements

This work was supported by National Natural Science Foundation of China(No:60773004) and scientific & technological key-project in Shanxi Province of China(No: 20090322004).

REFERENCES

[1] Simmons G J. The prisoners' problem and the subliminal channel [C]//Proceeding IEEE Workshop Communications Security CRYPTO'83, Santa Barbara, CA, 1983:51-67.

[2] Fridrich J, Soukal D.Matrix embedding for large payload [J] . IEEE Transactions on Information Security and Forensics, 2006, 1(3):390-394.

[3] Fridrich J,A Westfeld.High capacity despite better steganalysis (F5-a steganographic algorithm) [C]// Information Hiding, 4th International Workshop, Springer Verlag, New York. 2001:289-302.

[4] Ker A D.A capacity result for batch steganography [J].Signal Processing Letter, 2007, 14(8):525-528.

[5] Fridrich J, Lisonek P, Soukal D.On steganographic embedding efficiency [C]// Information Hiding 8th Inter-national Workshop, Alexandria, VA, USA, 2008:282-296.

[6] Hedieh S, Mansour J.Secure steganography based on embedding capacity [J].International Journal of Information Security. 2009.8: 433-445.

[7] Ker A D.Perturbation hiding and the batch steganography problem [C] // IH2008,Santa, USA,2008:45-59.

[8] Pevny T, Fridrich J.Benchmarking for steganography[C] // Information Hiding 2008. Santa Barbara, CA USA, 2008: 251-267.

[9] Ramkumar M, Akansu A N.Capacity estimates for data hiding in compressed images [J].IEEE Transactions on Image Processing, 2001, 10(8):1252-1263.

[10] Cachin C.An Information-theoretic model for stegano-graphy[C] // Inform ation Hiding 2nd International Workshop. Portland, USA,1998(1525): 306-318.

[11] Filler T, Ker A D.The square root law of steganographic capacity for markov covers[C]//Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia. San Jose, CA: 2009:18-21.

***Authors**: Asst.Prof. Ph.d.Gouxi Chen, North University of China , School of Electronics and Computer, 030051, Taiyuan, China, E-mail: chengouxicgx@163.com; M.A.Meng Zhang, North University of China , School of Electronics and Computer, 030051, Taiyuan, China, E-mail: zhanlangx0@gmail.com. Prof.Junjie Chen,* Taiyuan University of Technology,*E-mail:chenjunjie@tyut.edu.cn. Ph.d. Donglai Fu,E-mail: fudonglai@nuc.edu.cn;M.A.Yuliang E-mail:wuyuliang@163.com*