Steganalysis of LSB Speech with Low Embedding Rates based on Joint Probability

YANG Wanxia
School of Computer Science, China University of Geosciences, Wuhan430074, China
School of Mechatronical Engineering, Gansu Agricultural University, Lanzhou730070, China
15002612926 0086
E-mail:yangwanxia@163.com

TANGShanyu*
Information SecuritySchool of Computing and EngineeringUniversity of West London，
* Corresponding author: S. Tang
St Mary's Road, Ealing, London W5 5RF, UK
Tel: +44 (0)208 231 2948
E-mail:2412495341@qq.com

ABSTRACT

Steganalysis with low embedding rates is a difficult problem in the information hiding field. Using the characteristics of wavelet package decomposition, which is a delicate method for signal processing, and the feature of Joint Probability (JP), which better expresses the correlation of speech signals, a steganalysis at low embedding rates based on the JP feature of the second-order derivative-based Wavelet package Coefficient(WPC) of the speech signal is proposed in this paper. The steganography detection performance,including the detecting accuracy, computing complexity and extraction time of JP, was compared using Markov bidirectional transition probability (MBTP) and Traditional Markov second-order transition probability(TMSOTP) based on WPC and non-WPC, respectively. The experimental results indicate that the JP of the WPC is more suitable forsteganalysis at low embedding rates with the low complexity and shortest extraction time among the tested methods; the accuracy rate can be up to 75%,whereas the embedding is only 3%.

CCS Concepts

• CCS → Information systems → Information systems applications → Mobile information processing systems.

Keywords

Speech; Steganalysis; Joint probability; Markov transitionprobability.

1. INTRODUCTION

Because static carries cannot satisfy the real-time duplex interaction required by customers, a dynamic carrier of stream media such as Voice over IP (VoIP) is one of the most suitable carriers for secret communications[1].The International Telecommunication Union (ITU) has defined the high-rate speech codec G.711 for VoIP applications[2]. In addition, least significant bit(LSB) steganography can be efficiently applied to hide information in a waveform coder by directly modifying the least significant bits of each speech-sampling site. It is easy for streaming media software such as Skype and Gtalk to add steganography modules, which implies that secret communication channels can be built at will. Hence, the study of steganalysis on VoIP is vital and new.

However, the alteration of carrier structure caused by LSB steganography with low embedding rate is very small, and VoIP speech stream is real-time, it is hard to detect LSB steganography in VoIP. Research and practice results show that information hiding has changed inevitably the statistical characteristics of the cover, and Markov statistical model can depict the subtle change. SoMarkov statistical series methods are proposed to detect LSB steganography in VoIP and the detection performance is also compared between them.

The remainder of the paper is organized as follows. Section 2 introduces the related work on blindsteganalysis.Section 3 analyzes the second-order derivative-based spectrum of speech signal. Section 4 designs a method to extract the feature of the second-order derivative-based Wavelet package Coefficient(WPC) of speech signals. Section 5 describesthe Markov approach. Section 6 simulates and analyzes the experimental results. Section 7 concludes our work.

2. RELATED WORKS

Blind steganalysis focuses on extracting the effective features from large samples and selecting a suitable classifier. [3]selected the index vector of the reciprocal singular-value curve as the feature and was proven to be more accurate than SPAM(686D) and LogSV(25D) at different embedding rates. Other methods, [4,5]were proposed for low-embedding steganalysis. These algorithms emphasized the selection and optimization of features for steganalysis.

However, studies on steganography and steganalysis based on VoIP are comparatively insufficient because streaming media are structurally different from static carriers;However, the rapid growth of VoIP instant messaging has driven the development of steganography and steganalysis based on it. Huang et al. [6,7] improved the RS algorithm by introducing a sliding-window mechanism for adapting to VoIP speech real-time detection. Ren [8] selected the Markov transition probability of the adjacent scale factor band codebook as the feature and realized thesteganalysis of Hoffman codebook. Li et al. [9] realized the steganography detection of pitch modulationin compressed speech coding .Tian et al. [10] achieved the low-bit-rate detection of speech codec

ICCNS 2017, November 24–26, 2017, Tokyo, Japan
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5349-6/17/11…$15.00
https://doi.org/10.1145/3163058.3163067

fixed codebook index steganography by usingtheMarkov correlation of the pulse position.

These algorithms do not have ideal detection performance for steganography at low embedding rates. Thus, based on theory of derivative and wavelet packet decomposition of speech signal, the paper is aimed to achieve detection by using the principles of the Markov correlation statistical model. In consideration of the real-time detection of dynamic carrier, the paper mainly compares the detection performances of second-order derivative-based WPC and non-WPC high-order Joint probability(JP), Markov bidirectional transition probability(MBTP) and Traditional Markov second-order transition probability(TMSOTP) of speech signal at different embedding rates, particularlyat low embedding rates. In other words, the complexity of the feature calculation, time of feature extraction and discretion of the detectionaccuracy are not merely considered to improve the accuracy, In contrast to classic Mel-frequency cepstral coefficients(MFCC), the methods in this paper is superior.

Thesteganalysis methodproposed in this paperis shown in Fig. 1.

The methodology shown in Fig.1 is as follows:

1)speech signal samples are obtained;

2)the second-order derivative of samples obtained in step 1) is used;

3)samples in steps 2) are decomposed by wavelet packets, and the WPC are obtained;

4)the JP, MBTP and TMSOTP features of WPC and second-order derivative of samples in steps 2)are extracted; and

5) the SVM is trained by featuresin 4),and the experimental samples are classified.

3. SECOND-ORDER DERIVATIVE SPEDTRUM ANALYSIS

For any signal $f(t)\in L^2(R)$,according to the basic theory of signal processing, the Fourier transform of signal is

$$F(w)=\int_R f(w)e^{-jwx}dx \qquad (1)$$

Thus, its $n(n\in z^+)$ -order derivative is

$$D^n f(x)\overset{FT}{\Leftrightarrow}(DF)^n(w)=(jw)^n F(w)=\hat{d}^n(w)F(w) \qquad (2)$$

where $(jw)^n=\hat{d}^n(w)$ is the n-order differentialmultiplierfunction, i.e.,

$$\hat{d}^n(w)=\hat{b}^n(w)e^{j\theta^n(w)} \qquad (3)$$

In (3), $\hat{b}^a=|w|^a$, $\hat{\theta}^a(w)=\frac{a\pi}{2}\mathrm{sgn}(w)$, a is the order of derivative.

The amplitude value expresses the strength of the signal in the frequency domain in the formula $\hat{b}^a=|w|^a$,where the value of $|w|^a$ changes with a. On this basis, when a=0, 1, and2, the amplitude-frequency characteristic curve is experimentally drawn, as shown in Fig.2 .In the graph, the X-axis represents the frequency, and the Y-axis indicates the amplitude. The derivatives can significantly enhance the signalat high frequencies(a=2); therefore, it is widely used in fields such as signal processing. thesecond-order derivative is selected to analyze the speech signal, which is very important for speech steganalysis, particularly at low embedding rates.

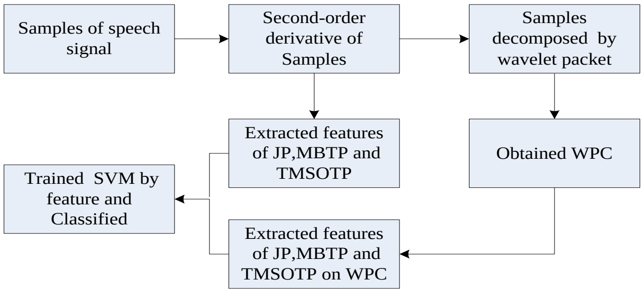
Samples of speech signal
Second-order derivative of Samples
Samples decomposed by wavelet packet
Extracted features of JP,MBTP and TMSOTP
Obtained WPC
Trained SVM by feature and Classified
Extracted features of JP,MBTP and TMSOTP on WPC

Figure. 1.Steganalysis method

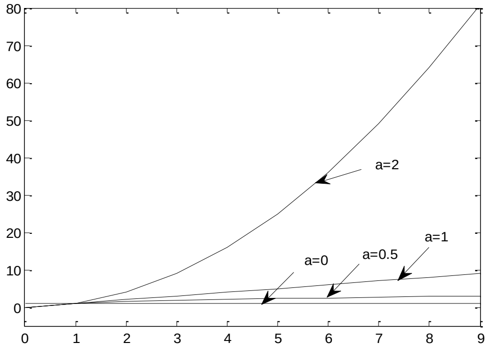

Figure2. Amplitude-frequency characteristic curve of a derivative-based speech signal

4. FEATURE EXTRATIONOF SECOND-DERIVATIVE WPC

Because the wavelet packet transform (WPT) has good time-frequency localization, the second-order derivative-based speech signals were decomposed by a wavelet packet to obtain more profiles and details of information.

Although speech signals are non-stationary, the wavelet packet decomposition performs the zooming function as follows:

$$\psi_{i,k}(\mathrm{t})=2^{-i/2}\psi(2^{-\mathrm{i}}\mathrm{t}-\mathrm{k}) \qquad (4)$$

where i is the scale index, k is the location index, and $2^{-i/2}$ is the magnification time. Smaller details of the signal can be acquired by reducing the i value. The WPT can zoom in on small details or brief outlines, so it can continuously decompose the low- and high-frequency signals.

Based on the theory, the original and steganography speech flows of the 16-bit mono PCM codec were decomposed by the db6 with the two-level wavelet packet;the second-layer approximate outlines and detail coefficients were also extracted from the samples. High-order JP, MBTP and TMSOTP were designed according to the obtained coefficients.

5. FEATURE EXTRACTION OF HIGH-ORDER JP,TMSOTPAND MBTP

5.1 Feature extraction of high-order JP

Because the speech signal is a certain correlation between similar parametersof adjacent frames.Thus, the sequence $S=\{s_1,s_2,\ldots\ldots,s_L\}$ of the inspected parameters can be considered the Markov chain.Currently, a high-order Markov transition probability matrix is widely used by many detection algorithms as a sensitive feature for speech steganalysis with higher accuracy.

However, the calculation amount will be tremendous and time-consuming when the dimensions of the high-order Markov transition probability increase, so the dimension-reduction measures are applied to make it suitable for the actual training and testing. Most dimension-reduction methods require a large amount of computation, which is not favorable for the real-time detection of VoIP speech streams. To this end, the frame number can be reduced, buta small number of detection frames will cause a very sparse matrix. Consequently, long samples are used to get a better classification effect, which increases the computation time.

Comparatively, a high-order joint probability is more capable of capturing the correlation of adjacent parameters,is less time-consuming and has a simpler algorithm.

The extraction processes of WPC and non-WPC JP features are as follows:

1)thesecond derivative is found for the speech signal;

2)thesecond-derivative-based speech signals are framed according to the short-time stationarity; each frame has 256 sampling points to avoid the overflow caused by overly high dimensions in calculating the high-order joint probability density;

3)each frame is decomposed by db6 with the two-layer wavelet packet; the contour and detail coefficients of the second-layer four sub-band are extracted;

4) the joint probability of the three adjacent coefficients of each wavelet sub-band is calculated; matrix summations for the joint probability are made with four wavelet sub-bands coefficients;and the matrix summations for the joint probability of each frame are computed, as shown in (5); and

5)for the comparative analysis, the aforementioned steps except 3) are used to extract the non-WPC high-order JPfeature ofthe second-derivative-based speech signal.

$$\mathrm{HJ}^{\mathrm{mn}}_{\lambda_1\lambda_2\lambda_3}=\sum_{k=1}^{n}\sum_{l=1}^{m}p(\mathrm{a}(\mathrm{i}+2)=\lambda_3,a(\mathrm{i}+1)=\lambda_2,a(\mathrm{i})=\lambda_1)\quad(5)$$

l is the number of sub-band; k is the number of frame; $p(a(i))=\lambda_i$ is the probability of the parameter; and $\mathrm{HJ}_{\lambda_1\lambda_2\lambda_3}$ is JP.

5.2 Extraction features of TMSOTP

The model of traditional second-order Markov is based on the hypothesis that the current speech signal is related to two preceding signals, as shown in (6).

$$MC^{\mathrm{mn}}_{\lambda_1\lambda_2\lambda_3}=\frac{\sum_{k=1}^{n}\sum_{l=1}^{m}p(\mathrm{a}(\mathrm{i}+2)=\lambda_3,a(\mathrm{i}+1)=\lambda_2,a(\mathrm{i})=\lambda_1)}{\sum_{k=1}^{n}\sum_{l=1}^{m}p(a(\mathrm{i}+1)=\lambda_2,a(\mathrm{i})=\lambda_1)}\quad(6)$$

l is the number of sub-band; k is the number of frame; $p(a(i))=\lambda_i$ is the probability of parameter; and $M\mathrm{C}_{\lambda_1\lambda_2\lambda_3}$ is TMSOTP. The processes are as follows:

1) the first three steps are identical tothose of the joint probability;

2)the sum of the transition probability matrix of four sub-and wavelet coefficients is computed; then, the sum of the transition probability matrix of each frame is calculated, as shown in (7); and

3)for thecomparative analysis, the aforementioned similar method except 3) in 5.1 is used to extract the non-WPC TMSOTPfeature of the second-derivative speech signal.

5.3 Extraction features of MBTP

The traditional second-order Markov has been improved;i.e., the joint probability of three adjacent coefficients is countedby taking the probability of the middle coefficients as the condition. The processes are as follows.

1)the first three steps are identical to those ofthe previous two methods;

2)inthe fourth step, matrix summations of the transition probability for the four wavelet sub-band coefficients are conducted; then, the summations of transition probability for each frame are performed as shown in (7).

$$MB^{\mathrm{mn}}_{\lambda_1\lambda_2\lambda_3}=\frac{\sum_{k=1}^{n}\sum_{l=1}^{m}p(\mathrm{a}(\mathrm{i}+2)=\lambda_3,a(\mathrm{i}+1)=\lambda_2,a(\mathrm{i})=\lambda_1)}{\sum_{k=1}^{n}\sum_{l=1}^{m}p(a(\mathrm{i}+1)=\lambda_2)}\quad(7)$$

l is the number of sub-band; k is the number of frame; $p(a(i))=\lambda_i$ is the probability of parameter; and $MB_{\lambda_1\lambda_2\lambda_3}$ is MBTP.

3)for the comparative analysis, the aforementioned method except 3) in 5.1 is used to extractthe non-WPC MBTPfeatures of the second-derivative speech signal.

6. TRAINING THE SVM CLASSIFIER AND RESULTANALYSIS

6.1 Training the SVM classifier

160 PCM speech segments of 16-bit codes were selected as the cover samples; each segment lasted 1 second at 8000Hz. In total, 160 Stego samples were obtained by LSB matching steganography at 1%, 3%, 5%, 8%, 10%, 30%, 50%, and 80% embedding rates. Of the 160 samples, 100 samples were trained, and the other 60 samples in each group were used for testing. LIBSVM3.1 was selected as the classifier set with default parameters andperformed in the following procedure:

1) 160 cover samples were selected for the LSB matching steganography, and stego samples were generated;

2)high-order JP features of WPC after the second derivative for 320 cover and stego samples were obtainedusing the method in 5.1, which formed the 62-dimension feature vector set L. Then, the non-WPC High-order JP features were extracted from 320 cover and stego samples of the second-derivative speech signal, which formed the 254-dimension feature vector set L1.

3)TMSOTP features of the WPC after the second derivative for 320 cover and stego samples were extracted using the method in 5.2, which formed the 62-dimension feature vector set M.

Then,the non-WPC TMSOTP features were extracted from 320 cover and stego samples of the second-derivative speech signal, which formed the 254-dimension feature vector set M1.

4)MBTP features of the WPC after the second derivative for 320 cover and stego samples were obtained using the method in 5.3, which formed the 62-dimension feature vector set N. Then, the non-WPC MBTP features were extracted from 320 cover and stego samples of the second-derivative speech signal, which formed the 254-dimension feature vector set N1.

5)The MFCC feature vectors were extracted from 320cover and stego samples for comparison with Markov feature.

6.2 Results analysis

Following the above procedures, the performance of LSB matching steganography detection was simulated using the feature sets L, M, N and L1, M1, N1. The detection accuracy of WPC and non-WPC features is shown in Tables 1 and 2, and the extraction time is shown in Tables 3 and 4.

In Tables 1 and 2, each Markov feature has better detection of LSB speech steganography than MFCC.However, JP as a feature is superior when the embedding rate is below 3%; the MBTP as a feature is obviously advantageous when the rate is above 5%; and the TMSOTP haspoor performance at all embedding rates. In contrast, the accuracy of the non-WPC feature is slightly higher than that of the WPC. A possible reason is that only the second layer of details and contour wavelet coefficients of the speech signal are extracted to compute the feature in this experiment, but the features of the non-WPC are calculated using the same speech signal, which contains relatively more information about the speech signal. Tables 3 and 4 show that JP as a feature has the shortest extraction time, followed by MBTP, and TMSOTP has the longest computation time. A comparison of Tables 3 and 4 shows that the extraction time of each WPC feature is much shorter than that of non-WPC features.

Table 1. Accuracy of the WPC features at various embedding rates

Embedding ratesFeature	1%	3%	5%	8%	10%	30%	50%	80%
JP	60%	75%	74.2%	78.3%	79.2%	83.3%	90%	91.7%
TMSOTP	60%	72.5%	72.5%	75.8%	79.2%	86.7%	86.7%	87.5%
MBTP	56%	69.5%	75%	78.9%	82.7%	89.7%	90.5%	95.9%

Table 2. Accuracy of the non-WPC features at various embedding rates

Embedding ratesFeature	1%	3%	5%	8%	10%	30%	50%	80%
NWJP	65%	75.9%	79.5%	81.5%	80.8%	82.5%	90%	95%
NWTMSOTP	63%	75%	80%	81.3%	83.2%	86.7%	91.7%	93.5%
NWMBTP	63%	75%	80%	82%	85%	90.5%	92.7%	97%
MFCC	58.3%	57.3%	65%	62.5%	62.5%	59.4%	56%	75%

Table 3. Extraction time of the WPC features at various embedding rates

Embedding ratesFeature	1%	3%	5%	8%	10%	30%	50%	80%
JP	1136s	1087s	957s	1009s	1041s	1376s	1662s	1941s
TMSOTP	1773s	1808s	1966s	2044s	1330s	1790s	2174s	2774s
MBTP	1432s	1674s	1338s	1558s	1289s	1573s	1791s	2186s

Table 4. Extraction time of the non-WPC features atvarious embedding rates

Embedding ratesFeature	1%	3%	5%	8%	10%	30%	50%	80%
NWJP	5448s	5341s	4768s	5147s	5249s	5491s	5689s	5978s
NWTMSOTP	6113s	6280s	6378s	6992s	6019s	6183s	7093s	7228s
NWMBTP	5776s	5879s	5608s	5690s	5593s	5771s	5864s	6005s

To more intuitively compare the detection accuracy and extraction time length of each feature, the detection accuracy of the three WPC features at3% and 30% embedding rates is shown in Fig. 3(a)(b). The detection accuracy of WPC and non-WPC JP features at 30% embedding rates is more intuitive in Fig. 3(c), and the extraction time of each feature at each embedding rate is shown in Fig.3 (d). The results in Fig. 3are consistent with Table 1 and 2.

7. CONCLUSION AND INNOVATION

Aiming todecrease the difficulty ofsteganography detectionat low embedding rates by fully using the second-order derivative-based speech signal WPC, which can more accurately describe the details of the signal, the experiment mainly compared the steganalysis performance of JP, MBTP and TMSOTP of WPC and non-WPC as features,which includes the detection accuracy, computation complexity and extraction time length. Large samples of speech flows were used to train the SVM classifiers. The results show that in terms of accuracy, with or without the wavelet packet decomposition, JP as a feature is dominant when the embedding rate is below 3%, whereas MBTP as a feature is obviously advantageous when the rate is above 5%. Comprehensively, the JP of WPC as a feature has the best performance, followed by MBTP of WPC, and TMSOTP of WPC has the worst performance.

Innovation point 1: the JP feature of WPC with the second derivative for speech signal is proposed to solvethe problem of LSB matching steganography with low embedding rate in high-rate speech coding.

Innovationpoint2:combining with the theory of digital signal processing,the proposed JP in this paper decreases the computation time and computational complexity and is more beneficial forreal-time detection

8. ACKNOWLEDGMENTS

This work was supported in part by grants from the National Natural Science Foundation of China (No.61402115, No.u1536115, No.u1536207). The authors would like to thank anonymous reviewers for their valuable suggestions.

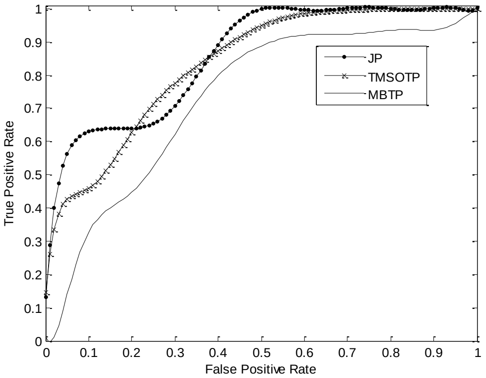

Figure3. (a) ROC curve of WPC features at 3% embedding rate

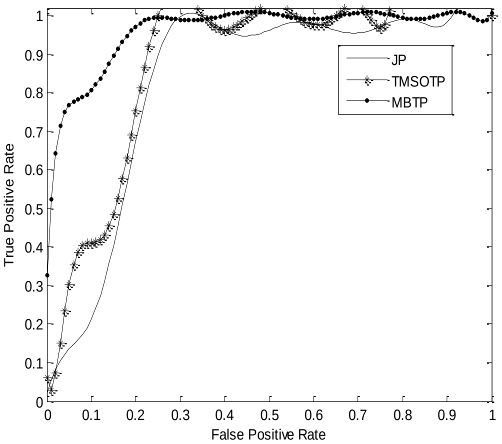

Figure3.(b) ROC curve WPC features at 30% embedding rate

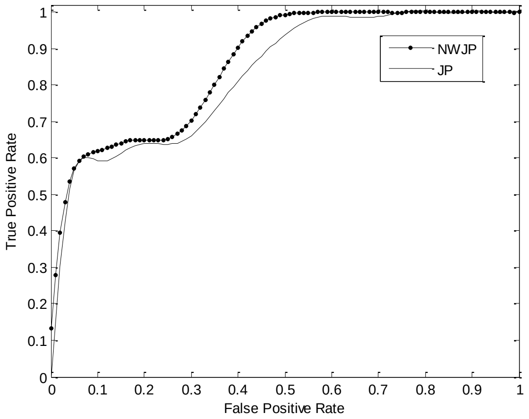

Figure3. (c) WPC and non-WPC JP features of ROC curve at 3%embedding rate

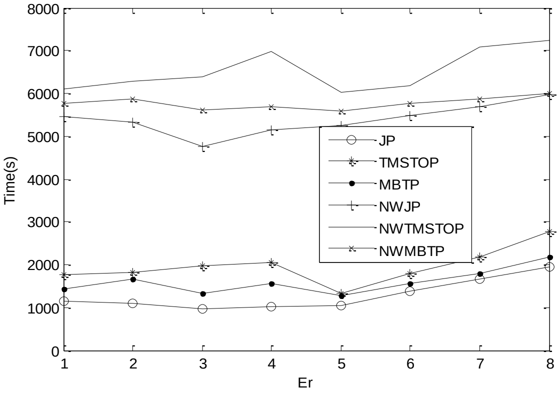

Figure3.(d)Timeconsumption comparison with each feature

9. REFERENCES

[1] Huang Yongfeng, Tang Shangyu.2016.Covert voice over internet protocol communications based on spatial model. SCIENCE CHINA(TechnologicalSciences).No.1,Vol.59(2016),117-128.

[2] Huang Yongfeng, LI SongBing. 2016. Network covert communication and its detection technology. Tsinghua University Press, Beijing.

[3] RoyaNouri,AzadehMansouri. 2015. Blind Image Steganalysis Based On Reciprocal Singular Value Curve. 9th Iranian Conference on Machine Vision and Image Processing.IEEE ,Tehran, Iran(Nov 2015),124-127.DOI= 10.1109/IranianMVIP.2015.7397519.

[4] SaeedAkhavan,Mohammad AliAkhaee.2015. SaeedSarreshtedari.: Images steganalysis using GARCH model for feature selection.J. Signal Processing. Image Communication. 39(Nov 2015),75-83.DOI=10.1016/j.image.2015.08.006.

[5] Geetha, S., Ishwarya, N. and Kamaraj, N. 2010.Audio steganalysis with Hausdorff distance higher order statistics

using a rule based decision tree paradigm.J. Expert Syst. Appl. 37(Dec 2010), 7469-7482. DOI=https://doi.org/10.1016/j.eswa.2010.04.012.

[6] Huang Y.F., Tang S.,Zhang, Y,Zhang.2011. Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis.J. IET Communications. 5(June2011),929-936.DOI= 10.1049/iet-com.2010.0348.

[7] Tao, H., Sun,D. andHuang,Y. 2014. A detection method of subliminal channel based on VoIP communication. In Proceedings of the 1st international workshop on Information hiding and its criteria for evaluation,.ACM, New York, NY, USA(June 2014),37-41.DOI=10.1145/2598908.2598910.

[8] YanzhenRen, QiaochuXiong, LinaWang. 2016.STEGANALYSIS OF AAC USING CALIBRATED MARKOV MODEL OF ADJACENT CODEBOOK. The 41st international conference on accoustic, speech and signal processing. IEEE,Shanghai, China(March 2016), 2139-2143.DOI= 10.1109/ICASSP.2016.7472055.

[9] LI Song-Bing,JIA Yi-Zhen, FU Jiang Yun, DAI Qiong-Xing.2014. Detection of pitch Modulation Information Hiding Based on Codebook Correlation Network .J. CHINESE JOURNAL OF COMPUTERS.VOL.37, No.10.(Oct 2014), 2107-2116.

[10] HuiTian,YanpengWu,Yongfeng Huang etal. 2015.Steganalysis of Low bit-rate Speech Based on Statistic Characteristics of Pulse Positions. 10th International Conference on Availability, Reliability and Security.IEEE, Toulouse, France (Aug 2015),455-460.DOI=10.1109/ARES.2015.21.