Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Audio steganalysis based on "negative resonance phenomenon" caused by steganographic tools*

RU Xue-min, ZHUANG Yue-ting‡, WU Fei
(*Institute of Artificial Intelligence, Zhejiang University, Hangzhou 310027, China*)
E-mail: ruxuemin@zju.edu.cn; yzhuang@cs.zju.edu.cn; wufei@zju.edu.cn
Received July 28, 2005; revision accepted Nov. 18, 2005

Abstract: Researching on the impact different steganographic software tools have audio statistical features, revealed the phenomenon that when messages are embedded in a WAV file by using a certain tool, the variation of statistical features in the WAV file which already contains messages embedded by the same tool is abruptly smaller than those in which messages have not been embedded. We call it "negative resonance phenomenon" temporarily. With the phenomenon above and Support Vector Machines (SVMs), we can detect the existence of hidden messages, and also identify the tools used to hide them. As shown by the experimental results, the proposed method can be very effectively used to detect hidden messages embedded by Hide4PGP, Stegowav and S-Tools4.

Key words: Audio steganalysis, Linear prediction, Support Vector Machine (SVM)
doi:10.1631/jzus.2006.A0577 **Document code:** A **CLC number:** TP309; TP391

INTRODUCTION

With the rapid progress in the Internet and expanded area of multimedia technology, steganography and steganalysis as subfields of information security have developed rapidly and drawn more and more attention. Steganography is the art of hiding and transmitting data through apparently innocuous carriers in an effort to conceal the existence of the secret data (Peticolas *et al*., 1999). While steganalysis, the countermeasure to steganography, is an art that detects even decodes hidden data within a given medium. Individuals or organizations may place personal/private/sensitive information in steganographic carriers (such as image, audio and video). However, on the contrary, steganographic techniques can be used for illegal activities committed by criminals or organized terrorists. Therefore steganalysis becomes the key technology to prevent the steganography from being used for illegal activities.

Steganalysis as a kind of novel information security technology evolved fast (Liang *et al*., 2004). In recent years, many steganalysis algorithms and methods have been emerging. Westfeld and Pfitzmann (1999) observed that when they were embedding encrypted data into a GIF image, the histogram of its color frequencies would be changed. They proposed χ^2 analysis on consecutive LSB embedding. Fridrich *et al*.(2001) presented RS (regular groups and singular groups) method that can be applied to 24-bit color images as well as to 8-bit grayscale (or color) images with randomly scattered message bits embedded in the LSBs of colors or pointers to the palette. Liu *et al*.(2004) proposed steganalysis of data hidden techniques in wavelet domain. Their approach lays particular stress on methods in which secret message was embedded via quantitating wavelet coefficients. Farid (2002) brought forward a universal method based on building higher-order statistical models for natural images and looking for deviations

‡ Corresponding author
* Project (No. 60272031) supported by the National Natural Science Foundation of China

from these models. Hiding data in audio files is a bit more challenging than hiding them in image files, as the human auditory system is more sensitive than its visual system. Compared to image steganalysis, audio steganalysis is relatively unexplored. Detection of steganographic modifications in an audio file can be made possible by testing its statistical properties (Johnson *et al*., 2005; Böhme and Westfeld, 2004). Deviation of statistical properties from a given norm can be identified as stego audio and is the key point for selecting the right statistical properties sensitive to the presence of an embedded steganographic message.

Our research focuses on WAV files. In order to discriminate stego audios from clear normal ones, we embed random data into a (possibly) stego WAV file by using a certain steganographic tool. It was found that the variation in some statistical features of WAV file is significantly different between clear WAV files and stego ones which already contain hidden messages embedded by the same tool. This phenomenon is similar to what we call resonance phenomenon in physics and know as an interesting phenomenon of the oscillation becoming stronger when the driving frequency matches the system's natural vibration frequency (its resonant frequency). But in this phenomenon, on the contrary, the variation in some statistical features of WAV file becomes smaller when the test tool matches the one used for inserting secret data into the stego WAV file. So we call it "negative resonance phenomenon" temporarily. With this phenomenon and Support Vector Machines, we can detect the existence of hidden messages, and also identify the tools used to hide them. As shown by the experimental results, the proposed method can be very effectively used to detect hidden messages embedded by Hide4PGP (Repp, 2005), Stegowav (Pulcini, 2005) and S-Tools4 (Brown, 2005).

The paper is organized as follows. In the next section, we describe the main ideas and the general methodology. Section 3 shows the experimental results of the proposed method. The last section will be devoted to discussing our implementation and outlining possible future directions of our research.

GENERAL METHODOLOGY

Our general methodology is to match the test steganographic tool with the one which may have been used in a (possibly) stego WAV file. The course of detection is composed of the following steps:

Step 1: Statistical features extraction. This step consists of 7 sub-steps.

(i) Read out the sample data S in the test WAV file. Then calculate the mean, variance, skewness and kurtosis of S.

(ii) Use linear predictor to get the error Err between S and its predicted value.

(iii) Calculate the mean, variance, skewness and kurtosis of Err.

(iv) Use a certain steganographic tool as the test tool, embed random message (random message can be obtained by compressing and encrypting a message) in the test WAV file at 100% steganographic capacity, and read out its sample data S'. Calculate the mean, variance, skewness and kurtosis of S'.

(v) Do the same as Step (ii) to get the error Err' between S' and its predicted value.

(vi) Do the same as Step (iii) to get the mean, variance, skewness and kurtosis of Err'.

(vii) Subtract the mean, variance, skewness and kurtosis of S' and Err' from the corresponding statistics of S and Err respectively. Use the results as input feature vectors of SVM.

Step 2: Non-linear rbf kernel SVM classifier training.

In this step, we collect 500 WAV files. Half of which are used as clear normal audios and the remaining as stego audios which are acquired by embedding random messages with the test steganographic tool at 60% steganographic capacity (for average condition). Then train a Support Vector Machine using the statistical features extracted from these 500 WAV files by following Step 1. This Support Vector Machine trained is the detector to be used as the test steganographic tool.

Step 3: Detection (classification).

Extract statistical feature from the (possibly) stego WAV files by following Step 1 and classify these files using the trained SVM in Step 2.

Hypothesis and principle

When people use steganography, they strive for high security and capacity. So it is common that the embedded message is compressed and encrypted. In fact, nowadays, compression and encryption are im-

plemented by steganographic tools themselves. Also, compressed and encrypted message is statistically random. Therefore we hypothesize that the hidden message is comprised of random data.

Steganographic techniques usually alter the statistics of the carrier and, obviously, longer hidden messages will alter the carrier more than shorter ones. For most steganographic techniques, it is usually not too difficult to identify a macroscopic quantity $F(m)$ that predictably changes (e.g., monotonically increases) with the length of the embedded secret message m. We call $F(m)$ the distinguishing statistics (Fridrich *et al*., 2003). Let us assume that the functional form of F is known or can be guessed from experiments. For example, F may be linear, quadratic, exponential, etc. In general, the function F will have some extreme values, such as $F(0)$ (for the cover audio) or $F(M_{\max})$ (for the stego audio with maximum message).

When we embed the maximum length random message in the (possibly) stego WAV file S, there are two cases:

If S is clear normal audio, the variation of the distinguishing statistics $F(m)$ is:

$$\Delta F_{\text{cover}}=F(M_{\max})-F(0). \tag{1}$$

If S already contains hidden message M_0, the variation of the distinguishing statistics $F(m)$ is:

$$\Delta F_{\text{stego}}=F(M_0+M_{\max})-F(M_0)\approx F(M_{\max})-F(M_0). \tag{2}$$

So

$$|\Delta F_{\text{cover}}|>|\Delta F_{\text{stego}}|. \tag{3}$$

However, among different steganographic tools, the distinguishing statistics and their variation regularities are different because of their different embedding algorithms. We can detect the existence of hidden messages, even identify the tools used to hide them, by examining the variation regularities of many statistical features jointly. Above is the theoretical foundation of the "negative resonance phenomenon". In order to clarify this phenomenon, consecutive embedding of 125 independent maximum length random messages in a clear WAV file was implemented using different steganographic tools. The variations of some statistical features are described in Fig.1.

In Fig.1, some statistical features vary abruptly at the first embedding, and these statistical features can be selected as distinguishing statistics. We can see that there may be several distinguishing statistics for a steganographic tool, and that the distinguishing statistics and its variation are not the same among different steganographic tools.

We select two sets of statistics as model statistics. One set is the mean, variance, skewness and kurtosis of audio S. These features provide information on the spatiotemporal distribution of the audio signal. The other set is from the errors which are the differences between S and its linear predictive values. These features provide information about the relationship among the neighbor audio samples.

Linear prediction

There are strong correlations between the consecutive samples of audio signal (especially speech signal) during a short-duration (about 20 ms). The current sample value of an audio signal can be estimated as a linear function of previous p samples.

In a frame with N samples, the most common representation is:

$$\hat{S}(n)=\sum_{i=1}^{p} a_i S(n-i), \tag{4}$$

where $\hat{S}(n)$ is the predicted signal value, $S(n-i)$ the previous values, and a_i the predictor coefficients. The prediction error is:

$$Err(n)=S(n)-\hat{S}(n)=S(n)-\sum_{i=1}^{p} a_i S(n-i), \tag{5}$$

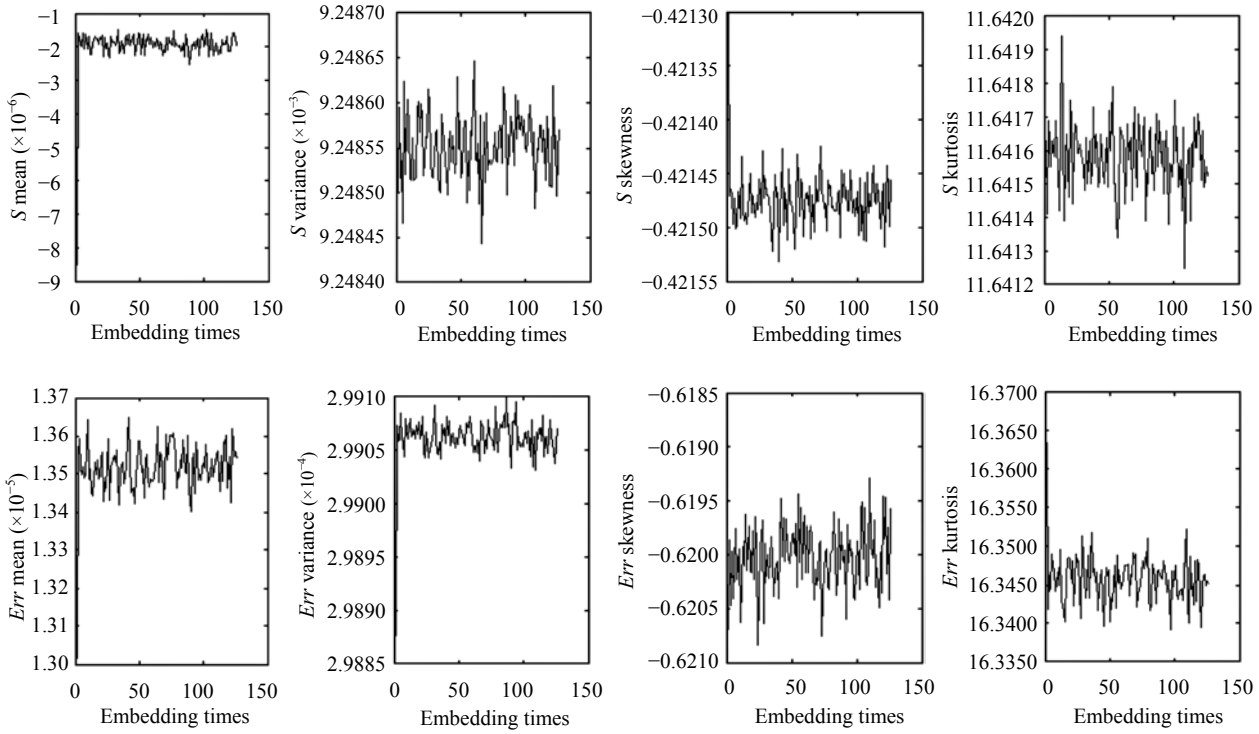
where $S(n)$ is the true signal value.

When dealing with the signals, we divide them into many short segments, called frames, which are usually 20~30 ms and overlapping. In order to get the most optimized prediction, we minimize the mean-squared prediction error. So the following set of linear equations must be solved:

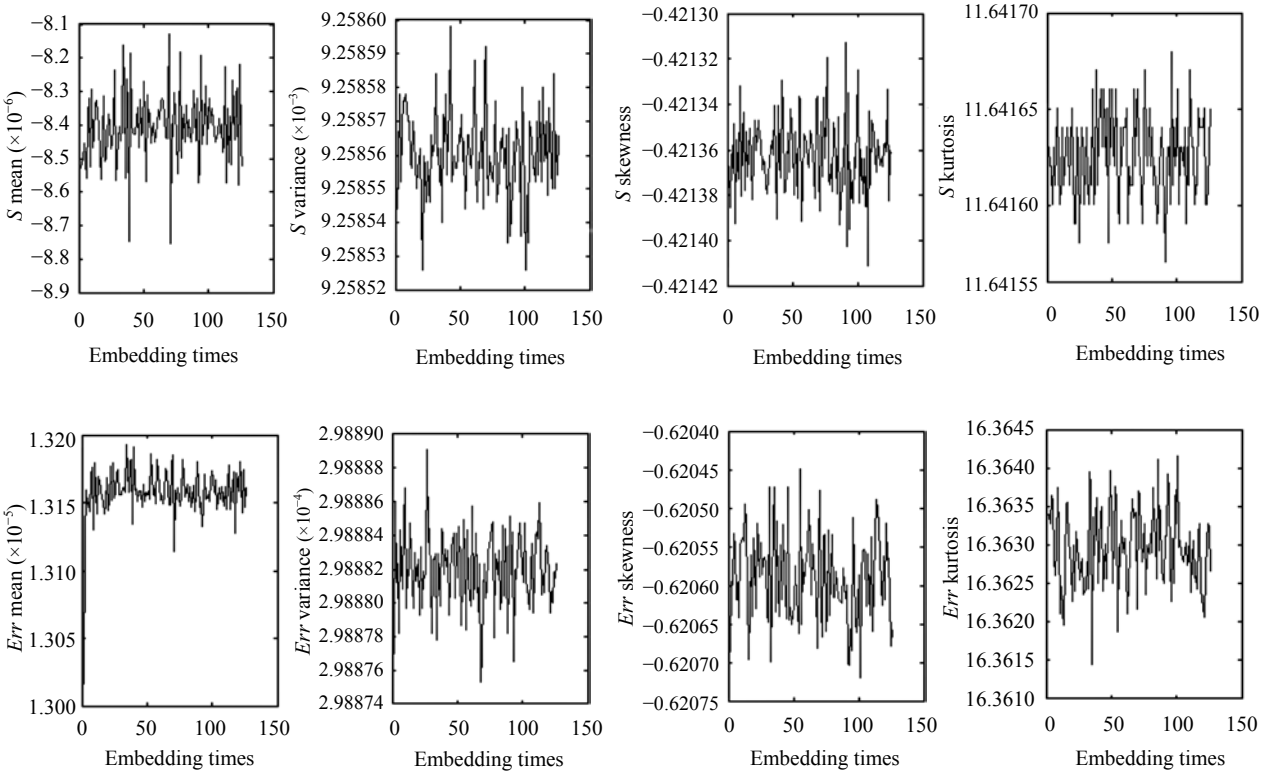
$$\sum_{k=1}^{p} a_k R(|i-k|)=R(i), \qquad 1\le i\le p, \tag{6}$$

where,

$$R(k)=\sum_{m=0}^{N-1-k} S(m)S(m+k). \tag{7}$$



(a)



(b)

Fig.1 The variations of some statistical features when we embed 125 random messages consecutively with maximum length in a clear WAV file (a) using Hide4PGP; (b) using Stegowav

In matrix form the set of linear equation can be expressed as:

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_p \end{bmatrix} = [R(1) \quad R(2) \quad R(3) \quad \dots \quad R(p)]^{\mathrm{T}}. \tag{8}$$

This particular system of equations can be solved by using the efficient Levinson-Durbin recursive procedure (Smith, 2004; Spanias, 1994). And then we get the predictor coefficients a_i and the prediction errors $Err(n)$.

As for our research work, the WAV file format is PCM with 44100 Hz sample rates, 16 bit depth and 2 channels. We analyze the audio signals under the frame of length N=512 samples with the frames overlapped by 50%. And let p=10, which means to estimate the current sample value with previous 10 samples.

Detection method

The process of distinguishing the audios with and without hidden data can naturally be viewed as a classification problem. And we refer to the categories as stego and normal. Support Vector Machine (SVM) classifier was used in our work because of its excellent performance (Vapnik, 1995; Burges, 1998). We used a set of audios (stego and normal audios) as the training data to build the SVM classifier. Given a set of audios, the SVM trained can be used for classification.

SVM techniques for classification are powerful tools for learning models that generalize well even in sparse, high dimensional settings. Based on Vapnik's statistical learning theory, SVM creates a maximum-margin hyperplane which separates the training vectors into different categories. When the margin is maximized, the probabilistic test error bound is minimized.

We used a non-linear rbf kernel SVM with 8 input features consisting of the differences between the mean, variance, skewness, kurtosis of S', Err' and the corresponding statistics of S and Err. Here the SVM trained is the detector of audio steganalysis.

EXPERIMENTAL ANALYSIS RESULTS

We tested our steganalysis technique on audio signals embedded with three steganographic tools: Hide4PGP v2.0, Stegowav and S-Tools4, which are available from the Internet.

Hide4PGP v2.0 is a command-line steganographic program for Windows, DOS, OS/2, and Linux that hides data within BMP, WAV, and VOC files. It is designed to be used with both PGP and Stealth, but also works well as a stand-alone program. Unlike other available steganography programs it has no own encryption. Version 2.0 has several new features, including a new stego format which is much more robust against format conversions—only lossy compression formats will lose the hidden data. Hide4PGP spreads the secret data evenly over the whole multimedia file. With real quantitative data it is able to use more than one bit per data point.

Stegowav (Pulcini, 2005) will hide a message into the LSBs (Least-Significant-Bits) of an RIFF file (8 or 16 bits per sample, PCM uncompressed) by changing the value of the LSBs for each sample to the value obtained from the message. The number of bits per sample that will be altered is set by the "-b #" switch. If the message is shorter than required to cover the full audio, random noise will be added.

S-Tools is a steganographic tool that hides files in BMP, GIF, and WAV files. When it hides data in sounds, S-Tools distribute the bit-pattern corresponding to the file you want to hide across the least significant bits of the sound sample. S-Tools seed a cryptographically strong pseudo-random number generator from your passphrase and use its output to choose the position of the next bit from the cover data to use.

Test results

We picked up 1000 Internet WAV (http://www.wavsurfer.com) files whose contents are from movies and television and transformed their compressed wave format into standard PCM wave format using Nero Wave Edit and divided them into two groups, 500 as normal audios. We made three copies of the other 500 WAV files. The Hide4PGP stego audios, Stegowav stego audios and S-Tools4 stego audios were created from the three copies by separately embedding messages at 60% steganographic capacity

with Hide4PGP, Stegowav and S-Tools4. The messages were compressed by WinRAR and encrypted by PGP beforehand.

For the purpose of illustration, consider first testing 500 normal audios and 500 Hide4PGP stego audios. "Negative resonance" test was implemented by embedding messages at 100% steganographic capacity and the set of statistical features was collected by following the steps described in Section 2. We added label −1 for statistical features from 500 normal audios and label +1 for statistical features from 500 stego audios and selected 250 statistical features labeled −1 and 250 statistical features labeled +1 to train a non-linear rbf kernel SVM. During our research, we used the LIBSVM2.6 (Chang and Lin, 2001) for our experiments. The left 250+250 statistical features were classified with the trained SVM.

Experiments on Stegowav and S-Tools4 were conducted in a similar way. The test results shown in Table 1 were satisfactory compared with other steganalysis techniques.

Table 1 Test results

	Hide4PGP	Stegowav	S-Tools4
Miss Det.	0/250	0/250	23/250
False Det.	16/250	7/250	49/250
Accuracy	96.8%	98.6%	85.6%

Cross test

We chose 80 WAV files from 1000 standard PCM format WAV files acquired previously, and divided them into four groups, 20 as normal audios, the remaining 60 included 20 Hide4PGP stego audios, 20 Stegowav stego audios and 20 S-Tools4 stego audios respectively embedded messages at 60% steganographic capacity with Hide4PGP, Stegowav and S-Tools4. The messages were compressed by WinRAR and encrypted by PGP beforehand.

The statistical features of 80 WAV files were collected by following the steps described in Section 2. We used the SVM model which had been trained aiming at Hide4PGP previously presented to classify the 80 WAV files. The test results are shown in Table 2.

Table 2 shows that the cross-correlation among different steganographic tools is weak although these three are all LSB methods. It is because of the different detail schemes. We guess that the cross-correlation may be noticeable only when the steganographic schemes are very similar.

Table 2 Cross test results

	Miss Det.	False Det.		Accuracy
Trained SVM model for Hide4PGP	0/20	Embedded by Stegowav	0/20	97.5%
		Embedded by S-Tools4	1/20	
		Normal audios	1/20	

DISCUSSION AND FURTHER WORK

In this paper, we have proposed a sort of novel audio steganalysis technique that is based on "negative resonance phenomenon" caused by steganographic tools. One can notice that this proposed scheme can not only detect the presence/absence of secret message, but also identify the tools used to hide it. This is very helpful for dictionary attack on suspecting files further. If we have the tool identified, perhaps the very tool can be used to extract the original message concealed.

In our experiment, the reason why we set the steganographic capacity to 60% is that the ratio of embedded file size to cover file size will typically affect the accuracy of just about any steganalytic technique (Farid, 2001) and this method is no exception. Thus, the key element to improve the sensitivity is to select the right statistical features.

With the research above, we will make an effort to examine more steganographic tools, find more significant statistical features, train correlated Support Vector Machine classifiers and build up an audio steganalysis system for known steganographic tools such as antivirus software for known computer virus.

References

Böhme, R., Westfeld, A., 2004. Statistical Characterisation of MP3 Encoders for Steganalysis. Proceedings of the Multimedia and Security Workshop. ACM Press, New York, p.25-34.

Brown, A., 2005. S-Tools4. Http://www.jjtc.com/stegoarchive/stego/softwarewindows.html.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**(2):121-167. [doi:10.1023/A:1009715923555]

Chang, C.C., Lin, C.J., 2001. LIBSVM: A Library for Support Vector Machines. Http://www.csie.ntu.edu.tw/~cjlin/

libsvm.

Farid, H., 2001. Detecting Steganographic Messages in Digital Images. Technical Report TR2001-412, Dartmouth College, Hanover, NH.

Farid, H., 2002. Detecting Hidden Messages Using Higher-order Statistical Models. Proceedings of International Conference on Image Processing. NY, **II**:905-908.

Fridrich, J., Goljan, M., Du, R., 2001. Reliable Detection of LSB Steganography in Color and Grayscale Images. Proceedings of the ACM Workshop on Multimedia and Security. Ottawa, CA, p.27-30.

Fridrich, J., Goljan, M., Hogea, D., Soukal, D., 2003. Quantitative steganalysis of digital images: estimating the secret message length. *ACM Multimedia Systems Journal, Special Issue on Multimedia Security*, **9**(3):288-302.

Johnson, M.K., Lyu, S., Farid, H., 2005. Steganalysis of Recorded Speech. Proceedings of Security, Steganography, and Watermarking of Multimedia Contents VII, SPIE San Jose, CA, USA, **5681**:664-672.

Liang, X.P., He, J.H., Li, J.Q., Huang, J.W., 2004. Steganalysis principle, actuality and prospect. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, **43**(6):93-96 (in Chinese).

Liu, S.H., Yao, H.X., Gao, W., 2004. Steganalysis of Data Hiding Techniques in Wavelet Domain. Proceeding of International Conference on Information Technology: Coding Computing. Las Vegas, USA, p.751-754.

Peticolas, F.A.P., Anderson, R.J., Kuhn, M.G., 1999. Information hiding—A survey. *Proceedings of the IEEE Special Issue on Protection of Multimedia Content*, **87**(7):1062-1078.

Pulcini, G., 2005. Stegowav. Http://www.jjtc.com/stegoarchive/stego/softwaredos.html.

Repp, H., 2005. Hide4PGP. Http://www.heinz-epp.online-home.de/Hide4PGP.htm.

Smith, J.O., 2004. Physical Audio Signal Processing. August 2004 Draft. Http://ccrma.stanford.edu/~jos/pasp/.

Spanias, A.S., 1994. Speech coding: a tutorial review. *Proceedings of the IEEE*, **82**(10):1541-1582. [doi:10.1109/5.326413]

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Spring-Verlag, New York, p.188-203.

Westfeld, A., Pfitzmann, A., 1999. Attacks on Steganographic Systems. Proceedings of Information Hiding—Third International Workshop. Springer-Verlag, Dresden, Germany, p.61-75.