



UWL REPOSITORY

repository.uwl.ac.uk

Identifying concealed channels in internet streaming media communication

Peng, Jinghui, Jiang, Yijing, Sofroniou, Anastasia and Tang, Shanyu ORCID logoORCID:
<https://orcid.org/0000-0002-2447-8135> (2026) Identifying concealed channels in internet streaming media communication. ACM Transactions on Internet Technology.

This is the Submitted Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/15488/>

Alternative formats: If you require this document in an alternative format, please contact:
open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

Identifying Concealed Channels in Internet Streaming Media Communication

Jinghui Peng

Yijing Jiang

Anastasia Sofroniou

Shanyu Tang

Due to the exponential increase in cyber/online crimes involving the use of steganography in streaming media, it is crucial to focus on studying steganalysis technology for the detection of covert communications. Several existing steganalysis methods rely on classifiers such as Support Vector Machines (SVM) and conventional statistical analysis.

However, these methods suffer from high computational intensity and time consumption, making them unsuitable for real-time detection of steganography in streaming media. Additionally, some steganalysis techniques are specifically designed for certain codecs or depend on particular steganographic algorithms, limiting their practicality and universality. To address these challenges, this study devises five steganalysis algorithms tailored for Voice over Internet Protocol (VoIP) to detect covert channels in streaming media communications. The algorithms are as follows: ANOVA, derivative-based Mel Frequency Cepstrum Coefficient (DMFCC), improved Regular and Singular (RS) test, Mann-Whitney-Wilcoxon (M-W-W) test, and FFT-based Steganalysis. Experimental evaluations were performed on two sets of media streams — one without steganography and the other with steganography. The results obtained from these experiments reveal that the proposed VoIP steganalysis algorithms exhibit variations in terms of detection sensitivity, whether blind or targeted, online or offline, as well as the requirement for original samples.

Keywords: ANOVA, Covert communications, DMFCC, Mann-Whitney-Wilcoxon, RS, Steganalysis, VoIP

1 INTRODUCTION

Steganalysis, a countermeasure against covert communication in the field of information hiding, is developing rapidly with the strong demand for an investigation into covert communication. Information hiding technology can be readily exploited by malicious actors, including hostile agents and terrorists. By hiding their secret information in streaming media carriers, they try to avoid content scrutiny and tracing, and use it to organise crimes and terrorist activities, steal military and/or commercial information, endanger national and public security, and undermine social stability. Steganalysis can also be used in healthcare information systems to offer adequate protection for streaming communications between healthcare devices. Besides, researchers studying steganography need appropriate steganalysis methods for their research. Therefore, it is imperative to develop information hiding detection technology for streaming media communications, especially steganalysis of streaming media.

According to the generalized information hiding model [1], information can be described as a perception space and a recording space. Assume that the full information space is n -dimensional, that is, $V^n = (v_1, v_2, \dots, v_n)$, one of the r_1 dimensional subspaces is a perceptual space, expressed as $W^{r_1} = (v_{s_1}, v_{s_2}, \dots, v_{s_r})$, one of the r_2 dimensional subspaces is a record space, expressed as $W^{r_2} = (v_{t_1}, v_{t_2}, \dots, v_{t_r})$. W^{r_1} and W^{r_2} do not completely overlap. Information hiding is the process of mapping the information in the perceptual space to a recordable but imperceptible space. The

legitimate receiver possesses the inverse mapping and key, allowing for correct mapping back to the perceptual space. Steganalysis aims to detect the imperceptible information in the case of unknown inverse mapping.

According to the comprehensive information theory [1], detection is defined as steganalysis at the grammatical information level. It involves using the feature code matching of the special mark generated by the steganalysis software to identify the carrier. Detecting differences in statistical features in carrier data, such as histogram analysis, chi-square analysis, RS analysis, and steganographic capacity analysis, falls under the category of steganalysis at the semantic information level. Steganalysis for multimedia carriers, including an analysis of the exchange of data between specific monitoring objects, a large amount of abnormal traffic, abnormal behavior and state of users, is called pragmatic information layer steganalysis, which can narrow the scope of specific analysis [1]. This study mainly focuses on steganalysis at the semantic level.

Unlike traditional image and audio steganalysis, steganalysis of streaming media generally starts from a pile of suspicious data streams (Transmitting messages are usually quite large.), and it is almost impossible to know whether data communication contains secret packets. There appear to be two modes here, which can be used to determine whether covert streaming media communication has taken place. One mode is offline steganalysis, which involves capturing and storing all the transmitted messages in a buffer, then using a very powerful computer for offline analysis. The other is online steganalysis,

which can detect in real time whether streaming data contains a secret message. Offline steganalysis (such as some existing machine learning methods) has an obvious disadvantage, since it requires a large buffer to store all data streams and consumes a significant amount of time. Therefore, this study focuses on online steganalysis, aiming to discover a fast and real-time method for network administrators to detect hidden secret messages in data streams.

The process of detecting whether secret data is hidden in the cover object is usually modelled as a classification problem. The steganalysis algorithm receives the objects to be tested as input, and classifies them into original cover objects and stego objects with data embedded. The steganalysis algorithms can be divided into two categories: Blind steganalysis and Targeted steganalysis. A steganalysis algorithm is described as blind if its method of detection is independent of the steganographic method used. Conversely, a targeted steganalysis method is one that is designed to detect the stego object that is created by one or more particular steganographic methods (e.g. LSB embedding). In both cases, steganalysis is modelled as a classification problem, and techniques available include pattern recognition, classification and machine learning. Generally speaking, machine learning requires more computational power and is more time-consuming than statistics. In blind methods, the object is usually represented in a high-dimensional feature space. Machine learning methods are then used to distinguish between the cluster of cover objects and the cluster of stego objects within the feature space. The important advantage

of blind methods is that they can potentially detect an unknown stego object. They are also capable of classifying the stego objects into different groups by using individual steganographic methods. As the targeted steganalytic methods need to be designed for each individual steganographic method, their implementation cannot be automated. On the other hand, such methods may be even more accurate than blind methods.

Furthermore, while deep neural networks represent a significant trend in steganography research, their practical deployment in real-time streaming media steganalysis faces considerable hurdles. These include demanding computational resources, the need for very large, labeled datasets for training, and often a lack of interpretability. This study consciously employs a portfolio of well-established statistical and feature-based methods. This strategic choice allows us to establish a transparent, computationally-efficient baseline and to critically evaluate the fundamental detectability of covert channels in VoIP, independent of the massive data and processing requirements of deep learning. The insights gained are crucial for guiding the future development of both lightweight and data-intensive steganographic solutions.

This comparative study provides a critical benchmark for researchers and practitioners in the field of information hiding and steganalysis. This paper presents a comparative study of five steganalysis algorithms for detecting covert channels in Voice over Internet Protocol (VoIP) communications: ANOVA, Derivative-based Mel Frequency Cepstrum Coefficient (DMFCC), Improved Regular and Singular (RS) test, Mann-Whitney-

Wilcoxon (M-W-W) test, and FFT-based Steganalysis. While numerous steganalysis methods exist, many are computationally intensive, codec-specific, or unsuitable for real-time application. This study evaluates these algorithms in terms of detection sensitivity, universality (blind vs. targeted), operational mode (online vs. offline), and dependency on original cover samples. The comparative analysis aims to identify which methods are most suitable for practical, real-time steganalysis in streaming media environments, thereby providing guidance for network administrators and security practitioners in selecting appropriate tools for covert communication detection. This empirical evidence is vital for guiding future research directions, such as the development of lightweight hybrid approaches or the optimization of machine learning models for streaming media. Ultimately, this research enables the security community to enact more effective and adaptive steganalysis strategies against the growing threat of covert channels in VoIP.

The rest of the paper is organised as follows: Section II describes the related work, Section III presents the five different steganalysis algorithms, Section IV depicts the experiments, including comparisons and discusses the results in detail. Finally, Section V concludes the paper.

2 RELATED WORK

There has been a sharp increase in cybercrimes leveraging streaming media [2]. While countermeasures such as Deepfake detection [3], image forgery detection [4], and watermarking [5-6] have advanced, research into streaming communication and audio

steganalysis has lagged behind, creating a critical gap that demands attention. Kraetzer and Dittmann first proposed a VoIP steganalysis method based on voice feature vector Mel-Cepstrum [7]. Assuming that secret information is not embedded fixedly and continuously into the voice carrier from the beginning to the end of the call, they used χ^2 statistical analysis to analyse the Mel feature of the hidden channel. Later, their work was extended to VoIP steganalysis using the Support Vector Machine (SVM) classifier [8]. Liu et al. also studied speech steganalysis based on time derivative spectrum and Mel eigenvector [9-10]. By calculating the distance between the speech signal and its denoising residue, Takahashi and Lee used a SVM classifier to detect hidden information [11]. This method could detect LSB, direct sequence spread spectrum (DSSS), frequency hopping spread spectrum (FHSS) and echo steganography. Garateguy et al. put forward a steganalysis method that relies on RTP packet classification and a random projection matrix [12]. It detected VoIP packets with hidden information through knowing a normal data stream structure, however, this method is computationally intensive and time-consuming. Arackaparambil et al. presented a statistical distribution-based steganalysis method for VoIP steganography [13], but the measurement range and the length of the detection window seriously affected the success rate of detection. For a low bit rate G.723.1 codec, Li and Huang suggested a voice stream steganalysis algorithm based on Complementary Neighbour Vertex - Quantization Index Modulation (CNV-QIM) [14], which uses a distributed histogram, a state transition model and a time-consuming SVM classifier. Li et

al. designed a mechanism for extracting the spread spectrum of hiding data from digital media by using multi-carrier/signature iterated generalised least squares (M-IGLS) [15], but it could only detect hiding images effectively. Janicki et al. studied the steganalysis method of VoIP transcoding information [16], but this method was only valid for the G.711 codec. Ren et al. stated that data hiding in AMR codec increased the probability of the same pulse position in the same track, and proposed a speech steganalysis algorithm based on a SVM classifier [17], but this method was effective only when the embedding rate was more than 30%.

Huang et al. proposed a steganalysis method for VoIP streaming media using a sliding window mechanism to improve the RS algorithm [18]. Compared with the RS algorithm, the detection time was reduced by 64%, but the detection experiment used the G.711 codec and VoIP data streams stored in a database, resulting in heavy computation. Additionally, for the G.723.1 codec, they also presented a steganalysis method based on quadratic statistics and regression analysis [19]. This method was used to analyse the compressed VoIP voice stream stored in the database and estimate the size of embedded secret information.

Tian et al. presented a tailored detection scheme for bitrate modulation-based steganography, which is based on the support-vector-machine [20]. And they developed SVM-based steganalysis for inactive speech frames, which employs the statistics for ZCC to characterise the frame-level dynamic characteristic of speech signals, and utilises the

average values of MFCCs to represent the invariant characteristic of inactive frames in 2020 [21]. Lin et al. designed a VoIP steganalysis algorithm called RNN-SM to detect QIM steganography. This codeword correlation model is based on a recurrent neural network (RNN), and a feature classification model was proposed to classify correlation features into cover speech and stego speech categories [22]. Yang et al. proposed a lightweight steganalysis scheme, which mapped vector quantization code-words into a semantic space and used the softmax classifier to categorize the input stream carriers [23]. And they introduced attention mechanisms to extract correlation in code-words for speech steganalysis tasks and designed a fast correlation extraction network without any RNN/CNN structure to detect QIM based steganography in VoIP streams [24]. Li et al. proposed a general frame-wise steganalysis method for low-bit-rate compressed speech, which used a dual-domain representation to conduct feature extraction both in the compressed domain and the decoded time domain and extract features from a speech frame [25].

The recent proliferation of deep learning in steganalysis has undoubtedly pushed the boundaries of detection accuracy. However, this advancement often comes at the cost of generality, interpretability, and practicality for resource-constrained, real-time scenarios like live VoIP monitoring. Many state-of-the-art deep learning models are architected for specific codecs or steganographic techniques, and their "black-box" nature makes it difficult to understand the root cause of detection failures or successes. In contrast, this

work returns to fundamental principles by re-evaluating and adapting classical statistical and signal processing techniques. Our goal is not to compete with deep learning on pure accuracy, but to provide a robust, efficient, and interpretable alternative for environments where deep models are impractical to deploy. This investigation into the core statistical properties disturbed by steganography provides valuable benchmarks and insights that can, in turn, inform the design of more efficient and targeted deep architectures in the future.

Steganalysis of streaming media steganography is far from mature. Some existing steganalysis methods need classifiers such as SVM and conventional statistical analysis, making them computationally intensive and time-consuming, thus failing to meet the requirements of real-time detection of streaming media steganography. Furthermore, other steganalysis methods are designed for specific codecs or rely on particular steganographic algorithms. This limitation makes them non-universal and difficult to implement in practical scenarios. Hence, there is a pressing need to explore a new steganalysis method for streaming media communications.

3 DETECTION ALGORITHMS OF COVERT CHANNELS IN VOIP

Steganographic communication is intended to conceal the existence of a steganographic message hidden in innocuous transmissions such as VoIP on the network. Statistical undetectability is normally used to evaluate the security of a steganographic system [26], i.e. a measure of how difficult it is to determine reliably the existence of a secret message hidden in a cover object. In other words, a secure steganographic system means a

statistically undetectable system. In this study, five steganalysis algorithms, i.e., ANOVA, Mel Frequency Cepstrum Coefficient (DMFCC), Improved Regular and Singular (RS) test and Mann-Whitney-Wilcoxon (M-W-W) test, are devised to detect covert channels in VoIP communications.

3.1 ANOVA test

ANOVA stands for analysis of variance and is used to test for differences in means between two or more groups. It is often used to determine if there is a significant difference between groups based on a categorical variable. ANOVA can be used to test for differences between groups, and to determine if there is an interaction between groups.

ANOVA has three main assumptions: (1) Homogeneity of variance; (2) The population follows normal distribution; (3) Each sampling is independent.

The formula for calculating ANOVA is:

$$F = (MS_{between}/MS_{within}) \quad (1)$$

where:

$MS_{between}$ is the mean square between groups, and MS_{within} is the mean square within groups.

3.2 Derivative-based Mel Frequency Cepstrum Coefficient (DMFCC)

Through close analysis of VoIP streaming media signals, it is known that the characteristic parameters of VoIP audio signals include pitch period, linear prediction parameter (LPC),

line spectrum pair parameter (LSP), Cepstrum Coefficient, Perceptual Linear Predictive parameters (PLP) and WV spectrum parameters. The main challenge in pitch detection lies in accurately identifying each pitch period along with its start and end positions within voice segments. The pitch period varies from person to person and has a wide range of variation. Linear predictive parameters are widely used in the fields of speech recognition, speech coding and speech synthesis. Lattice method proposed by Makhoul is particularly effective. Line spectrum pair parameter is another form of linear prediction parameter, which is closely related to the spectral characteristics of speech signals, and its quantisation and interpolation characteristics are superior to reflection coefficients. Perceptual Linear Predictive parameters are based on the characteristic parameters of the auditory model, which are easily interfered by the frequency response of the communication channel. WV spectrum is a high-order, non-stationary signal, which involves complex time and frequency domain transformation, and is related to the choice of frame length. There are three kinds of cepstrum parameters for VoIP audio signals: LPCC parameters, MFCC parameters, and ASCC parameters.

Mel frequency cepstrum represents a linear transformation of the logarithmic energy spectrum based on the nonlinear Mel scale of sound frequency. The four stages to the calculation of Mel frequency cepstrum are as follows: decompose a speech signal into multiple speech segments, pre-emphasize the speech signal, remove the interference of low-frequency signals through a high pass filter, and then perform Fourier transform on

the speech signal. The Mel scale is obtained by passing the spectrum obtained from each speech segment through the Mel filter. Then it calculates the logarithmic energy on each Mel scale. The above results are then transformed into the cepstrum domain by inverse discrete Fourier transform. The Mel Frequency Cepstrum Coefficient (MFCC) represents the magnitude of the cepstrum.

Due to the auditory characteristics of human ears, the Mel Frequency Cepstrum Coefficient (MFCC) first converts a speech spectrum into a non-linear spectrum based on Mel frequency standard, and then converts it into the cepstrum domain. MFCC has no premises / assumptions, good speech recognition ability and the ability to handle noise, with high computational complexity and accuracy. Preliminary research shows that MFCC has better performance than LPCC in language recognition. The filter set is used to calculate Mel cepstrum coefficients. Since the human perception of sound above 1000 Hz follows an approximate linear relationship in the logarithmic frequency coordinates, these filters have equal width in the Mel frequency coordinates. The Mel cepstrum coefficient $C_{mel}(n)$ is affected by various factors, such as the number, shape, distribution and energy spectrum of the filters in the filter set. It can be obtained through a modified inverse discrete cosine transform (IDCT):

$$C_{mel}(n) = \sum_{k=1}^K \theta(M_k) \cos\left(n(k - 0.5)\frac{\pi}{K}\right) \quad (n = 1, 2, \dots, p) \quad (2)$$

In the formula, $\theta(M_k)$ denotes the output energy of the k th filter, and p is the order of MFCC parameters.

The most widely used steganalysis method based on Mel spectrum, DMFCC, is the method in Liu et al. [27]. A speech steganalysis method based on spectrum analysis and Mel cepstrum analysis of second derivative was proposed. In spectrum analysis, the statistical characteristics of the high-frequency spectrum of the second derivative are extracted as spectral features. The Mel cepstrum coefficient derived from the second derivative was taken as the characteristic parameter. In addition, in order to compare with the second derivative method, a method based on wavelet spectrum and Mel cepstrum was also designed, and a support vector machine based on RBF (radial basis function) kernel was used to detect and distinguish steganographic audio signals. The results showed that the method based on reciprocal and wavelet was better than Kraetzer and Dittmann's method, and the success rate of detection was the highest when MFCC based on second derivative was used as the characteristic coefficient.

Support vector machine (SVM) is a machine learning classifier. The goal is to automatically classify the given data once learning. SVM classifier mainly includes three processes: training process, testing process and prediction process. The training speech samples include the original speech and the speech with embedded secret information. In the training process, the features that are sensitive to information hiding and conducive to the classification of the classifier are extracted from a large number of training speech

samples. In DMFCC method, the feature is MFCC parameters, which are used to construct the classifier of steganalysis and train the classifier. During testing, known data samples are used to evaluate the trained model's performance and generalization capability. In the prediction phase, features are extracted from the test speech samples. The extracted feature parameters are consistent with the features of the training classifier in the training process. The classifier utilises the extracted features for classification. There are three kinds of kernel functions commonly used in SVM classifiers, polynomial kernel function, radial basis function RBF kernel function and sigmoid kernel function. Among them, RBF-based SVM has good performance and learning ability in practical applications.

3.3 Improved Regular and Singular (RS) test

The RS (regular and singular) detection algorithm, originally proposed in [28], was designed specifically for detecting LSB-based image steganography. The RS image detection process mainly includes three stages: simulating embedding, revealing existence of steganography, and estimating the embedding rate. First, embedding simulation is the process of modifying the lowest bit of the pixel. If the original pixel value is an even number $2k$, change 0 to 1, then modify it to $2k + 1$, and the pixel value increases; If the original pixel value is an odd number $2k+1$, modify the lowest bit to change 1 to 0, then modify it to $2k$, and the pixel value decreases. The equation F is constructed to represent the hidden data, that is, to reflect the change between $2k + 1$ and $2k$. Modifying the

lowest bit of the pixel will not change the pixel value $2k$ to $2k - 1$ or $2k + 1$ to $2k + 2$.

This situation is expressed as the F' equation.

The main idea of RS detection is as follows:

- (1) LSB steganography is equivalent to performing the F operation on some pixels;
- (2) If F and F' operations are performed on some pixels of the original carrier, such as the pixel group G , statistically, the degree of chaos of the image block is the same. The discrimination function is used as the measurement function $f(G)$ of the degree of chaos, and the original image satisfies $f(F(G)) = f(F'(G))$;
- (3) After LSB steganography, the original carrier G becomes G' , and then applies F' , its chaos degree has $f(F'(G')) > f(F(G'))$.

(4) The asymmetry of chaos indicates the existence of secret information. The chaos degree discriminant function f is defined as: $f(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n-1} |X_{i+1} - X_i|$.

According to changes in the statistical characteristics of the pixel values of the image before and after LSB steganography, the original pixel group G is divided into three categories by the chaos measurement function f and operation F . $f(G)$ is the chaos degree before f operation, and $f(F(G))$ is the chaos degree after f operation:

Regular group: when $f(F(G)) > f(G)$, $G \in R$

Singular group: when $f(F(G)) < f(G)$, $G \in S$

Unusable group: when $f(F(G)) = f(G)$, $G \in U$

If vector M is used to represent different embedding types, F_M represents different LSB hiding methods, M is of length n (the same length as the pixel group G), and the elements are any combination of 0 and 1, with elements 0 and 1 belonging to the uniform distribution.

R_M and S_M represent the numbers of regular groups and single groups obtained by the carrier after F_M operation, respectively. When there is no information embedding, the values of R_M and R_{-M} are approximately the same, and the values of S_M and S_{-M} are also roughly the same. After LSB random steganography, $R_M \approx R_{-M}$ and $S_M \approx S_{-M}$ cannot be satisfied, and the degree of chaos is asymmetric.

Fridrich et al. [28] proved through a large number of experiments and calculations that R_M , S_M , R_{-M} and S_{-M} meet the following relationships:

$$2(d_1 + d_0)x^2 + (d_{-0} - d_{-1} - d_1 - 3d_0)x + d_0 - d_{-0} = 0 \quad (3)$$

where $d_0 = R_M(\frac{p}{2}) - S_M(\frac{p}{2})$, $d_1 = R_M(1 - \frac{p}{2}) - S_M(1 - \frac{p}{2})$, $d_{-0} = R_{-M}(\frac{p}{2}) - S_{-M}(\frac{p}{2})$, $d_{-1} = R_{-M}(1 - \frac{p}{2}) - S_{-M}(1 - \frac{p}{2})$.

$R_M(\frac{p}{2})$, $S_M(\frac{p}{2})$, $R_{-M}(\frac{p}{2})$, $S_{-M}(\frac{p}{2})$ can be calculated directly from the image to be measured, and then the LSB bits of all pixel values are replaced by 0 and 1, and then $R_M(1 - \frac{p}{2})$, $S_M(1 - \frac{p}{2})$, $R_{-M}(1 - \frac{p}{2})$ and $S_{-M}(1 - \frac{p}{2})$ can be calculated. The embedding

rate of secret information $p=x/(x-0.5)$ can be estimated by calculating the value of x through formula (3).

However, the experiment showed that the p obtained by using the formula in the original carrier without any secret information is not equal to 0, and the p value at this time is called the initial deviation. Assuming that the initial deviation of the original carrier is p_0 , the embedding detection rate calculated after LSB embedding is p_{det} , and the final secret information embedding rate $p = \frac{p_{det}-p_0}{1-p_0}$ can be obtained by correcting the embedding rate.

Huang et al. also set the threshold value V in their study [29] to determine whether there is information hiding. When the estimated $p \geq V$, there is hidden information; When the estimated $p < V$, there is no hidden information. In order to reduce the detection time, the computational complexity and the storage space, and realize the real-time detection of VoIP, the embedding length can be estimated after the calculation of $R_M(\frac{p}{2})$ and $R_{-M}(\frac{p}{2})$, as follows [29]:

$$\theta_r = \frac{R_M\left(\frac{p}{2}\right) - R_{-M}\left(\frac{p}{2}\right)}{R_M\left(\frac{p}{2}\right) + R_{-M}\left(\frac{p}{2}\right)} \quad (4)$$

Assuming V_t as the optimal detection threshold, redefine V_t . When $\theta_r < V_t$ in formula (4), it means that there is no hidden information; Otherwise, the secret information is hidden.

RS detection first needs to set M and the threshold V , where the number of 0s and 1s in M is related to the size of the group. If there are n samples in each group, there are n elements of 0 and 1 in M . In summary, the RS detection process is summarized as follows:

(1) Divide the speech G to be tested into k groups, namely $G = \{G_1, G_2, G_i \dots G_k\}$, with n samples in each group, and calculate the original chaos degree $f(G_i)$ for each group.

The chaos degree function is: $f(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n-1} |X_{i+1} - X_i|$, where X_i represents the value of each sample point.

(2) Perform F_M and F_{-M} transformations on each G_i in G to obtain the corresponding $2k$ $F(G_i)$, where M is a 0, 1 combination, $-M$ is a 0, -1 combination, and the corresponding elements of M and $-M$ are opposite. Then calculate the chaos degree $f(F(G_i))$ for each $F(G_i)$.

(3) According to the rule: when $f(F(G)) > f(G)$, $G \in R$; When $f(F(G)) < f(G)$, $G \in S$; When $f(F(G)) = f(G)$, $G \in U$. Calculate R_M, S_M, R_{-M} and S_{-M} . R_M represents the number of regular groups obtained by the carrier after F_M operation, S_M represents the number of singular groups obtained by the carrier after F_M operation, R_{-M} represents the number of regular groups obtained by the carrier after F_{-M} operation, and S_{-M} represents the number of singular groups obtained by the carrier after F_{-M} operation, that is, $R_M(\frac{p}{2})$, $S_M(\frac{p}{2})$, $R_{-M}(\frac{p}{2})$, and $S_{-M}(\frac{p}{2})$.

- (4) Modify all the lowest bits of the original G_i , that is, LSB bits, and all the lowest bits 0 are changed to 1, 1 is changed to 0, and G_i' is obtained. Then, perform steps 2 and 3 on G_i' to obtain $R_M(1 - \frac{p}{2})$, $R_{-M}(1 - \frac{p}{2})$, $S_{-M}(1 - \frac{p}{2})$, and $S_M(1 - \frac{p}{2})$.
- (5) The value of p can be calculated by using the formula (3) or the improved formula (4), where θ_r is the embedding rate.
- (6) According to the calculated embedding rate p and the set threshold V , judge whether the original voice contains any secret information. The judgment rules are as follows: if $p > V$ or $\theta_r > V$, it indicates hidden information in the voice carrier; otherwise, there is no hidden information.

3.4 Mann-Whitney-Wilcoxon (M-W-W) test

Instead of normal statistical analysis, the Mann-Whitney-Wilcoxon (M-W-W) test was used in this study to perform statistical undetectability analysis to evaluate the security of a steganographic algorithm. The test is one of the best-known non-parametric significance tests that assess whether two independent samples of observations come from the same distribution [30]. A comparison of the probability distributions between the original and stego VoIP streams samples reveals whether the differences in the probability distributions are almost indistinguishable.

According to the principles of the M-W-W and statistics, when the sample sizes are sufficiently large (at least 12 each), the test is based on the standardized test statistic [30].

In an examination of the steganographic algorithm for the real-time steganographic system with embedded VoIP, the test statistic can be computed as follows:

First, combine the n_1 sample of observations from Population 1 (*i.e.* the original VoIP streams samples) and the n_2 sample of observations from Population 2 (*i.e.* the stego VoIP samples with the hidden message), and arrange the combined data in ascending order;

Secondly, assign ranks to the combined observations (starting with 1 for the smallest observation);

Finally, sum the ranks of the n_2 sample of observations from Population 2, and denote this sum by S_2 .

The test statistic z^* is given by

$$z^* = \frac{S_2 - E\{S_2\}}{\sigma\{S_2\}} \quad (5)$$

where $E\{S_2\}$ and $\sigma\{S_2\}$ are the mean and the square root of the variance of the sampling distribution S_2 , respectively, which is the combination of the two samples of observations to be assessed.

4 RESULTS AND DISCUSSION

In our experiments, steganalysis tests on VoIP communications were carried out on original VoIP streams and stego VoIP Streams to be detected using the five steganalysis algorithms detailed in the previous section, respectively. The original and stego VoIP streams used in each test were obtained from the experiments of our earlier study

published in [31]. To ensure the reproducibility and reliability of our experiments, all tests were conducted on a standardized hardware and software platform. The experimental machine was equipped with an Intel Core i7-10700K CPU @ 3.80GHz, 32 GB of RAM, and ran the Windows 10 operating system. The algorithms were implemented and executed in MATLAB R2013a and Excel, with the exception of the DMFCC method which utilized the LIBSVM library for SVM classification. In the communication system, VoIP audio samples are coded with single-channel PCM having a sampling rate of 8000 Hz, which are used as cover objects for steganography. Each speech sample is represented with 16 bits. The different embedding methods employed in the covert VoIP communication system were based on the data embedding algorithms with variable embedding capacities, which were achieved by varying embedding location intervals[31].

4.1 ANOVA

The original and stego VoIP stream samples were tested by using ANOVA in this study. The sizes of samples (data points) were 4096 and 4096 in computation of the test. Our null hypothesis is that there is no difference in the original and stego VoIP stream samples. Our alternative hypothesis is that there is a significant difference.

Step 1: Calculate the mean and variance of each sample.

Step 2: Calculate the grand mean.

Step 3: Calculate the between-group sum of squares (SSB) and the within-group sum of squares (SSW).

Step 4: Calculate the degrees of freedom (df) for the F-distribution.

Step 5: Calculate the mean square values (MSB and MSW).

Step 6: Calculate the F-statistic.

Step 7: Determine the p-value and compare it to the significance level.

Among the variables shown in Table 1, the P-value is 0.954383, which is greater than 0.05. It fails to reject the null hypothesis, indicating no significant difference between the tested samples. The high P-value indicates that the variance between the original and stego groups is statistically indistinguishable from the variance within the groups. The results show that ANOVA did not detect the existence of hidden messages. This highlights the limitation of basic parametric tests against advanced, adaptive steganography.

Table 1: Undetectability analysis results using ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.60E-06	1	2.60E-06	0.003272	0.954383	3.842954
Within Groups	6.498324	8190	0.000793			
Total	6.498326	8190				

4.2 Derivative-based Mel Frequency Cepstrum Coefficient (DMFCC)

A software package called LIBSVM can be used to realize the function of SVM. It can quickly and effectively carry out SVM pattern recognition and regression. It not only provides the execution file that can be directly used in Windows systems, but also provides the source code for improvement.

In order to realise the DMFCC method, the MFCC parameters of speech samples were extracted first, and then a SVM support vector machine classifier was used to train and classify a large number of speech samples without and with hidden secret information. In this study, the above process was implemented in Matlab2013. That is, the DMFCC method utilized an SVM classifier to train a large number of speech samples, and then the trained classifier was used to test the speech and obtain the test results.

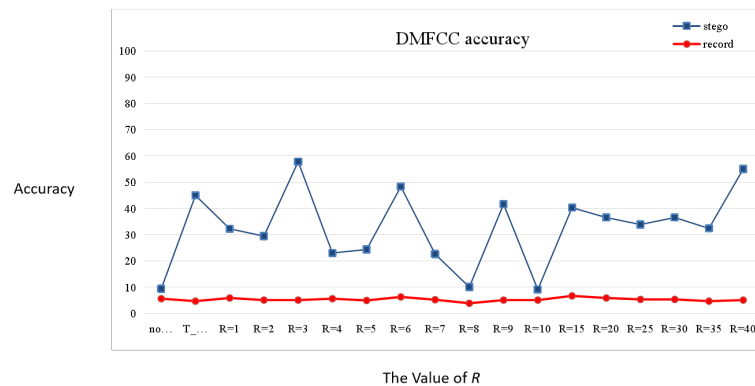


Fig. 1. DMFCC test results

In the DMFCC method, LIBSVM was used to train and test speech classification. The speech containing secret information is marked as -1, and the speech without hidden secret information is marked as 1.

As shown in Figure 1, stego represents the speech containing embedded secret information. Assuming that it is marked as -1, the accuracy of predicting -1 (i.e., the probability of correct detection) is less than 60%, indicating that the speech does not contain secret information. Assuming it is marked as 1, the accuracy of prediction of 1 is less than 10%. Parameter R is the interval to embed the bit stream of secret message into every byte of audio[31]. It can be seen that the performance of the DMFCC detection algorithm is varying, and the detection accuracy is relatively low.

The low classification accuracy of the DMFCC-SVM model (below 60% for stego speech) indicates that the Mel-cepstral features extracted from the derivative spectrum are not sufficiently discriminative for this specific steganographic method. A potential reason is that the adaptive embedding of introduces perturbations that are smaller than the natural variance of MFCCs across different speech segments, causing the SVM classifier to fail in finding a robust decision boundary in the feature space.

4.3 Improved Regular and Singular (RS) test

RS detection obtains the embedding rate of secret information in the voice file through statistical analysis of all voice samples, that is, the ratio of the secret information bitstream that replaces the least significant bit of voice samples. In Table 2, ‘stego’ represents the voice file containing secret information, ‘record’ represents the voice file without secret information embedded, and P_d represents the embedding rate directly detected by RS; P is an improved method for calculating the embedding rate, which is given by $P = (p_d - p_0)/(1 - p_0)$, where p_0 represents the initial deviation, and the minimum deviation value p_0 of 0.65 is determined based on a large number of experimental results; Q is the calculation result of the improved embedding rate in formula (4), that is, $Q = (R1 - R2)/(R1 + R2)$. From Table 3, it can be observed that for the record voice files, the original embedding rate P_d , the improved embedding rate P and the Q obtained by RS detection do not differ significantly from the results of stego voice files, indicating that the steganographic algorithm [31] is capable of resisting RS detection. The near-identical values of P and Q for both stego and record samples demonstrate that the improved RS test cannot reliably distinguish them. This implies that the embedding process does not introduce the statistical asymmetry in local sample sequences that the RS test is designed to detect, showcasing its robustness against this form of steganalysis.

Table 2: RS test results

Speech samples	Stego $(P=(p_0-p_0)/(1-p_0), p_0=0.65)$			Record $(Q=(R_1-R_2)/(R_1+R_2))$		
	P_d	P	Q	P_d	P	Q
no embedding	0.666666667	0.047619048	0	0.667790922	0.050831207	-0.005025126
T LSB	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=1	0.667790922	0.050831207	-0.00502513	0.666666667	0.047619048	0
R=2	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=3	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=4	0.666666667	0.047619048	0	0.66554619	0.044417686	0.005025126
R=5	0.666666667	0.047619048	0	0.66554619	0.044417686	0.005025126
R=6	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=7	0.667790922	0.050831207	-0.00502513	0.666666667	0.047619048	0
R=8	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=9	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=10	0.666666667	0.047619048	0	0.66554619	0.044417686	0.005025126
R=15	0.664406545	0.041161557	0.01010101	0.666666667	0.047619048	0
R=20	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=25	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=30	0.666666667	0.047619048	0	0.66554619	0.044417686	0.005025126
R=35	0.666666667	0.047619048	0	0.666666667	0.047619048	0
R=40	0.666666667	0.047619048	0	0.666666667	0.047619048	0

4.4 Mann-Whitney-Wilcoxon (M-W-W) test

To improve the precision of statistical analysis of undetectability, the sizes of the samples (data points) were 401 and 401 in computation of the test. Table 1 shows the test results and the parameters used to compare the probability distribution drawn from the original VoIP streams samples to that drawn from the stego VoIP stream samples that embed the secret message steganographically.

As shown in Table 3, the values of the test statistic z^* for five tests were much smaller than 1.960, which is the threshold, *i.e.*, $|z^*| \leq 1.960$, the results were H_0 (meaning two distributions do not differ) at all the tests, indicating that the original and stego VoIP stream samples do not differ. The results suggest that the M-W-W test cannot detect the

existence of the secret message embedded in the original VoIP stream using the steganographic algorithm [31].

Table 3: Undetectability analysis results using M-W-W test

Test	Sample size, n_1	Sample size, n_2	Sum, S_2	Mean of variance, $E\{S_2\}$	Variance, $\sigma^2\{S_2\}$	Test statistic, z^*	H
Test 1	401	401	160054	161001.5	10760267	-0.2888	H_0
Test 2	401	401	159060	161001.5	10760267	-0.5919	H_0
Test 3	401	401	160273	161001.5	10760267	-0.2221	H_0
Test 4	401	401	161297	161001.5	10760267	0.09008	H_0
Test 5	401	401	159035	161001.5	10760267	-0.5995	H_0

4.5 FFT-based Steganalysis

In our previous study, a steganalysis method based on Fast Fourier Transform (FFT) was devised to detect the covert channel of mobile streaming media in real time [32]. FFT is a method used to convert data from the time domain to the frequency domain. By comparing the distribution differences of two groups of audio streams in the frequency domain, it can recognise whether they are different and then determine whether the audio streams to be detected contain any secret data.

In the experiments, steganalysis tests on VoIP communications were conducted on original VoIP streams and stego VoIP Streams that were to be detected. To detect whether the the VoIP streams contain secret data, the distributions of the two groups of original and stego VoIP streams were compared in the frequency domain.

The experimental results suggested there was a significant difference between the power spectra in the frequency domain for the original VoIP streams and the stego VoIP stream that were to be detected, indicating that the VoIP streams that were to be detected contain hidden data and FFT-based steganalysis is effective in detecting covert channels in VoIP communications.

4.6 Comparison of proposed steganalysis algorithms

Five proposed steganalysis algorithms were compared in this study. Table 4 shows a comparison of detection sensitivity, universality, online/offline and need for original samples between ANOVA, Mel Frequency Cepstrum Coefficient (DMFCC), improved Regular and Singular (RS) test, Mann-Whitney-Wilcoxon (M-W-W) test and FFT-based Steganalysis [32]. Close analysis shows these five detection algorithms have different sensitivity. All the algorithms except for the Regular and Singular test are of blind steganalysis. DMFCC is an offline test, and others are online. DMFCC and improved RS tests do not need original samples for detection, whereas the other three algorithms require them.

Table 4: Comparison of proposed steganalysis algorithms

Algorithm	Detection Sensitivity	Universality	online/offline	Whether original samples needed
ANOVA	Low	Blind	Online	Yes
DMFCC	Moderate	Blind	Offline	No
Improved RS	Low	Targeted	Online	No
M-W-W	Low	Blind	Online	Yes
FFT	High	Blind	Online	Yes

The assertion that many existing methods are computationally intensive primarily stems from their dependence on complex SVM models with high-dimensional feature spaces. In contrast, the algorithms evaluated in this work offers a path toward real-time capability. Specifically, the ANOVA, Improved RS, M-W-W, and FFT-based methods are inherently based on lightweight statistical and signal processing operations. This avoids the costly training and feature extraction phases of machine learning models, making them suitable for online analysis as confirmed in Table 4. While the DMFCC method employs an SVM and is classified as offline, its inclusion provides a critical performance benchmark, illustrating the trade-off between the higher detection sensitivity of complex models and the operational requirement for low-latency, online analysis. Therefore, this work demonstrates that for scenarios demanding real-time performance, lightweight statistical and transform-domain methods present a viable and efficient alternative.

A key advantage of our selected algorithm portfolio is its movement away from the limitations of codec-specific or steganography-specific detection. As summarized in Table

4, four of the five proposed algorithms (ANOVA, DMFCC, M-W-W, and FFT-based analysis) are blind steganalysis methods. Their detection principles are based on general statistical discrepancies or feature anomalies, rather than a particular embedding technique. This makes them broadly applicable across different VoIP codecs and resilient against unknown steganographic methods. The inclusion of the Improved RS test, a targeted method, serves a specific purpose: to evaluate the resilience of modern steganography against well-known detection paradigms. The fact that it failed to detect the covert channels (as shown in Table 2) underscores the sophistication of current steganographic algorithms and reinforces the need for the more general, blind approaches that form the core of our comparative study.

5 CONCLUSION

In this study, five different steganalysis algorithms for the detection of covert communication over streaming media were proposed. The results of the experiments suggest that the proposed steganalysis algorithms differ in detection sensitivity, whether they are blind or targeted, whether they are online or offline, and whether they require original samples. Considering detection sensitivity, the FFT-based steganalysis method and the Mel Frequency Cepstrum Coefficient (DMFCC) method are highly capable of detecting the existence of the secret message embedded in VoIP streams using the steganographic algorithm [31], but the DMFCC method can only be used offline.

Further studies are therefore necessary to investigate how to accurately acquire the media packets needed for detection when heavy packet loss occurs.

ACKNOWLEDGMENTS

This work was supported in part by the Education Department of Guangdong Province under Grant 2021KTSCX063, Special topic of basic and applied basic research in Guangzhou under Grant SL2023A04J01043, Guangdong Regional Joint Fund under Grant 2022A1515110693 and GPNU Science Foundation under Grant 2021SDKYA 025.

REFERENCES

- [1] X. Niu, Y. Yang, Study on the Frame of Information Steganography and Steganalysis, ACTA Electronica Sinica, 34.B12(2006):4.
- [2] J. Lubacz, W. Mazurczyk, and K. Szczypiorski (2010). Voice over IP. IEEE Spectrum, 40-45.
- [3] X. Liao, Y. Wang, T. Wang, J. Hu and X. Wu (2023). FAMM: Facial Muscle Motions for Detecting Compressed Deepfake Videos Over Social Networks, IEEE Transactions on Circuits and Systems for Video Technology, 33(12), 7236-7251.
- [4] J. Chen, X. Liao, W. Wang, Z. Qian, Z. Qin and Y. Wang (2023). SNIS: A Signal Noise Separation-Based Network for Post-Processed Image Forgery Detection, IEEE Transactions on Circuits and Systems for Video Technology, 33(2), 935-951.
- [5] L. Fu, X. Liao, J. Guo, L. Dong and Z. Qin (2025). WaveRecovery: Screen-Shooting

- Watermarking Based on Wavelet and Recovery, IEEE Transactions on Circuits and Systems for Video Technology, 35(4), pp. 3603-3618.
- [6] Y. Li, X. Liao and X. Wu (2024). Screen-Shooting Resistant Watermarking With Grayscale Deviation Simulation, IEEE Transactions on Multimedia, vol. 26, 10908-10923.
- [7] C. Kraetzer, and J. Dittmann (Jan. 2007). Mel-Cepstrum based steganalysis for VoIP-steganography. Proc. IS&T/SPIE Symposium on Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII, San Jose, USA, pp. 650-661.
- [8] C. Kraetzer, and J. Dittmann (2007). Pros and cons of Mel-Cepstrum based audio steganalysis using SVM classification. Information Hiding, Springer Berlin Heidelberg, 359-377.
- [9] Q. Liu, A. H. Sung, and M. Qiao (2009). Temporal derivative-based spectrum and mel-cepstrum audio steganalysis. IEEE Transactions on Information Forensics and Security, 4(3), 359-368.
- [10] Q. Liu, A. H. Sung, and M. Qiao (2011). Derivative-based audio steganalysis. ACM Transactions on Multimedia Computing, Communications and Applications, 7(3), 1-19.
- [11] T. Takahashi T, and W. Lee (2007). An assessment of VoIP covert channel threats. Proc. 3rd Int Conf Security and Privacy in Communication Networks, Nice, France,

371-380.

- [12] G. Garateguy, G. Arce, and J. Pelaez (2011). Covert channel detection in VoIP streams. Proc. 45th Annual Conference on Information Sciences and Systems, Baltimore, Maryland, 1-6.
- [13] C. Arackaparambil, G. Yan, S. Bratus, and A. Caglayan (2012). On tuning the knobs of distribution-based methods for detecting VoIP covert channels. Proc. Hawaii International Conference on System Sciences, Hawaii, 2431-2440.
- [14] S. Li, H. Tao, and Y. Huang (2012). Detection of quantization index modulation steganography in G. 723.1 bit stream based on quantization index sequence analysis. Journal of Zhejiang University Science C, 13(8), 624-634.
- [15] M. Li, M. K. Kulhandjian, D. A. Pados, S. N. Batalama, and M. J. Medley (2013). Extracting spread-spectrum hidden data from digital media. IEEE Transactions on Information Forensics and Security, 8(7) 1201-1210.
- [16] A. Janicki, W. Mazurczyk, and K. Szczypiorski (2014). Steganalysis of transcoding steganography. Ann. Telecommun., vol. 69, 449-460.
- [17] Y. Ren, T. Cai, M. Tang, and L. Wang (2015) AMR steganalysis based on the probability of same pulse position. IEEE Transactions on Information Forensics and Security, 10(9), 1801-1811.
- [18] Y. F. Huang, S. Tang, and Y. Zhang (2011). Detection of covert Voice over Internet Protocol communications using sliding window-based steganalysis. IET

Communications, vol. 5, iss. 7, 929-936.

- [19] Y. Huang, S. Tang, C. Bao, and Y. J. Yip (2011). Steganalysis of compressed speech to detect covert Voice over Internet Protocol channels. *IET Information Security*, 5(1), 26-32.
- [20] H. Tian, J. Sun, C. -C. Chang, Y. Huang and Y. Chen, "Detecting Bitrate Modulation-Based Covert Voice-Over-IP Communication," in *IEEE Communications Letters*, vol. 22, no. 6, pp. 1196-1199, June 2018, doi: 10.1109/LCOMM.2018.2822804.
- [21] H. Tian, J. Liu, C. -C. Chang, Y. Huang and Y. Cai, "Detecting Steganography in Inactive Voice-Over-IP Frames Based on Statistic Characteristics of Fundamental Frequency," in *IEEE Access*, vol. 8, pp. 6117-6129, 2020, doi: 10.1109/ACCESS.2019.2962009.
- [22] Z. Lin, Y. Huang and J. Wang, "RNN-SM: Fast Steganalysis of VoIP Streams Using Recurrent Neural Network," in *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1854-1868, July 2018, doi: 10.1109/TIFS.2018.2806741.
- [23] H. Yang, Z. Yang, Y. Bao, S. Liu and Y. Huang, "Fast Steganalysis Method for VoIP Streams," in *IEEE Signal Processing Letters*, vol. 27, pp. 286-290, 2020, doi: 10.1109/LSP.2019.2961610.
- [24] H. Yang, Z. Yang, Y. Bao, S. Liu and Y. Huang, "FCEM: A Novel Fast Correlation Extract Model For Real Time Steganalysis Of VoIP Stream Via Multi-

- Head Attention," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 2822-2826, doi: 10.1109/ICASSP40776.2020.9054361.
- [25] S. Li, J. Wang and P. Liu, "General Frame-Wise Steganalysis of Compressed Speech Based on Dual-Domain Representation and Intra-Frame Correlation Leaching," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2025-2035, 2022, doi: 10.1109/TASLP.2022.3182276.
- [26] J. Fridrich, (2014). Steganography in digital media: principles, algorithms, and applications, 2nd ed. Cambridge University Press.
- [27] Q. Liu, A. H. Sung, M. Qiao. Temporal derivative-based spectrum and mel-cepstrum audio steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2009,4(3):359-368.
- [28] J. Fridrich, M. Goljan, R. Du. Detecting LSB steganography in color, and gray-scale images[J]. IEEE Multimedia, 2001, 8(4): 22-28.
- [29] Y. Huang, S. Tang, Y. Zhang. Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis[J]. IET COMMUNICATIONS, 2011,5(7):929-936.
- [30] J. Neter, W. Wasserman, & G. A. Whitmore, (1993). Applied statistics. 4th ed. Simon & Schuster.
- [31] S. Tang, Y. Jiang, L. Zhang, Z. Zhou, Audio steganography with AES for real-

time covert voice over Internet protocol communications, Science China Information Sciences. 57 (3)(2014) 1-14.

- [32] J. Peng, S. Tang , and J. Li . Fast Fourier Transform-Based Steganalysis of Covert Communications over Streaming Media. World Academy of Science, Engineering and Technology (WASET) 7(2019).