State-based steganography in low bit rate speech

Conference Paper · October 2012

DOI: 10.1145/2393347.2396395

4 authors, including:

Ke Zhou
Huazhong University of Science and Technology
146 PUBLICATIONS **980** CITATIONS
SEE PROFILE

Hui Tian
National Huaqiao University
91 PUBLICATIONS **834** CITATIONS
SEE PROFILE

Chunhua Li
Huazhong University of Science and Technology
14 PUBLICATIONS **77** CITATIONS
SEE PROFILE

State-based Steganography in Low Bit Rate Speech

Ke Zhou and Jin Liu
School of Computer, Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, 430074, China
raidkick@263.net, geneleocn@gmail.com

Hui Tian
College of Computer Science and Technology
National Huaqiao University
Xiamen 361021, China
cshtian@gmail.com

Chunhua Li *
School of Computer, Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology
Wuhan, 430074, China
li.chunhua@hust.edu.cn

ABSTRACT

Common least significant bit (LSB) relevant steganography methods in speech frames often base on bits evaluation by certain speech quality evaluation criterion and together with some coding and embedding strategies to enhance imperceptibility and efficiency. However, some embedding capabilities and security strategy are neglected. This paper proposes a state-based steganography method which fully investigates speech frame features in order to expand embedding capabilities and enhance steganography security. In the proposed method, embedding capabilities are measured by available numbers of states relative to current frame parameters rather than information bits. And secret information embedding procedure is performed by the chosen states mapped operations. This is a basic fine-grained steganography solution which is useful for other algorithms based on it. The experimental results have demonstrated that state-based method outperforms traditional LSB substitution in overall performance and introduces limited latency, which meets the realtime requirement in covert communications.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection, Data communications*; K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Insurance, Invasive software, Unauthorized access*

General Terms

Algorithms, Design, Experimentation, Security

Keywords

Steganography, State-based, Least Significant Bit, Low bit rate speech

1. INTRODUCTION

Steganography is the art of hiding secret messages over an innocent cover without knowing the existence of hiding. This emerging technique gives opportunity to covert communication but it also raises concerns about security and privacy problems [3]. The underlying steganographic covers are always some kinds of digital media such as images, documents or audio. With the advancement of computer communication networks, steganography has also penetrated into Internet based protocols and payloads [15], which brings both opportunities and challenges to communication security. In both of these two types of hiding applications, least significant steganography (LSB) related algorithms are always most popular and effective [10]. When it comes to low bit rate speech which is widely used as the payload in real time communications such as voice over Internet protocol (VoIP), restricted bandwidth and severe latency requirement are the most limiting factors for steganography [12].

In VoIP environment, a large amount of redundant data can be employed for steganography. They are distributed in protocol headers (e.g. IP, UDP/TCP, RTP/RTCP, SIP) [1, 15] and payloads (compressed pcm data or speech frames) [9, 2]. Lubacz et al. [5] classified VoIP specified steganography into three types which include a third type implemented by modifying time related fields of VoIP packets [4]. However, due to the existence of active wardens and limited hiding capacity, protocol based algorithm are generally for synchronization and supplementary usage [11]. Therefore, the payload based steganography always bear the main information of covert communication [6]. In [14] Xiao et al. proposed a quantization index modulation (QIM) based algorithm which utilized the redundancies of quantization codebooks in parameter speech codecs. In the method only 3 bits/frame were obtained, which is not sufficient. Huang et al. [2] referred to a silent frame hiding method with embedding rate up to 101 bits/frame. Both the active and inactive frame hiding are based on LSB substitution method which still has a capability promotion. Some stego coding schemes like matrix coding [13] used to enhance embedding efficiency or imperceptibility always base on it.

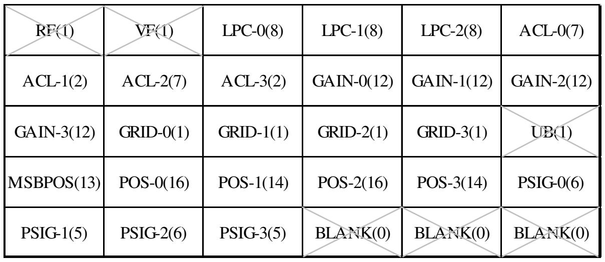
This paper proposes a novel state-based steganography approach which expands the LSB method to enhance embedding capabilities. Those capabilities can be used to improve the embedding efficiency, imperceptibility and embedding security. In the proposed method, we choose widely used ITU-T G.723.1 [7] codec for evaluation, on which two kinds of embedding operations corresponding to two state-based strategies are performed to get the best performance. With the evaluated optimal parameter states, intra-parameter and inter-parameter based key security schemes are available to increase the confidentiality of embedded private information.

The rest of this paper is organized as follows. Section 2 specifies the proposed state-based method and its embedding and extracting procedures. Then perceptual evaluation and security analysis of two state-based strategies are described in Section 3.1. Finally, Section 4 concludes this paper.

2. STATE-BASED METHOD

2.1 G.723.1 Speech Codec

ITU-T G.723.1 codec is designed for speech compression at a very low bit rate [7]. The codec contains two bit rates: 5.3 and 6.3 kbit/s. The latter is chosen for our proposed method, which has greater speech quality. G.723.1 belongs to parameter codecs and can be similarly considered as a two dimensional table consisting of corresponding frame parameters. Figure 1 depicts the 27 frame parameters of G.723.1 in a 5×6 table (3 BLANK padding), where the numbers in parentheses represent number of bits in each parameter. In the frame the first two parameters representing silent frame and frame bit rate, as well as the UB (unused bit) parameter which is vulnerable to steganalyzers, should not be modified. Thus, the rest 24 parameters are under consideration for steganography.



RF(1)	VF(1)	LPC-0(8)	LPC-1(8)	LPC-2(8)	ACL-0(7)
ACL-1(2)	ACL-2(7)	ACL-3(2)	GAIN-0(12)	GAIN-1(12)	GAIN-2(12)
GAIN-3(12)	GRID-0(1)	GRID-1(1)	GRID-2(1)	GRID-3(1)	UB(1)
MSBPOS(13)	POS-0(16)	POS-1(14)	POS-2(16)	POS-3(14)	PSIG-0(6)
PSIG-1(5)	PSIG-2(6)	PSIG-3(5)	BLANK(0)	BLANK(0)	BLANK(0)

Figure 1: G.723.1 6.3kbit/s frame parameter

2.2 Parameter States Strategies

According to Figure 1, a G.723.1 speech frame can be denoted as $\mathcal{F} = \{p_i \mid 0 \leq i \leq 26, i \in \mathbb{Z}\}$, where p_i is the ith parameter of the frame. When $\mathcal{F}$ is utilized for hiding, $i \notin \{0, 1, 17\}$. And $(p_i)_2 = b^i_{n_i-1} \cdots b^i_k \cdots b^i_1 b^i_0$ denotes the binary format of p_i, where n_i is the number of bits in $(p_i)_2$ and b^i_k is the kth bit of it. $(p'_i)_2 = {b'}^i_{n_i-1} \cdots {b'}^i_k \cdots {b'}^i_1 {b'}^i_0$ denotes the stego cover (speech cover after embedding), where ${b'}^i_k$ is either b^i_k or $\tilde{b}^i_k$. Here, $\tilde{b}^i_k$ is the negation of b^i_k.

In LSB based method, prior processed secret message M is embedded in the form of bits. Each bit contains two basic states: 0 and 1. The embedding function of p_i is a map from cover bit to stego bit $f : c \rightarrow s$, where $c, s = \{0, 1\}$. The performed $0 \rightarrow 1$ or $1 \rightarrow 0$ bit translation on b^i_k can be regarded as $(p_i - a) mod\, 2^{n_i}$ or $(p_i + a) mod\, 2^{n_i}$ arithmetical operation, where $a = |p'_i - p_i|$. Let $s^i_0, s^i_1, \ldots, s^i_{2^{n_i}-1}$ denote 2^{n_i} different states of p_i. Then one-bit LSB embedding is equivalent to a map $f' : c' \rightarrow s'$, where $c', s' = \{s^i_0, s^i_1, \ldots, s^i_{2^{n_i}-1}\}$ and $(p_i - a) mod\, 2^{n_i}, (p_i + a) mod\, 2^{n_i} \in s'$. Table 1 depicts two kinds of state-based strategies, where $(p_i - 2^{n_i-1}) \equiv (p_i + 2^{n_i-1}) mod\, 2^{n_i}$ in arithmetic based strategy. The two state-based strategies both have totally 2^{n_i} states.

Table 1: State-based Strategies

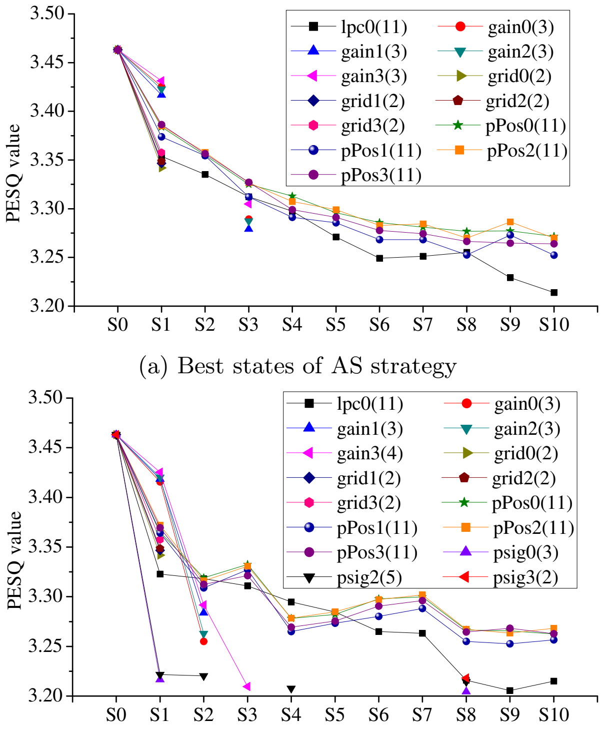
State	Arithmetic Based (AS)	Bits Based (BS)
s^i_0	$p_i mod\, 2^{n_i}$	$b^i_{n_i-1} \cdots b^i_1 b^i_0$
s^i_1	$(p_i + 1) mod\, 2^{n_i}$	$b^i_{n_i-1} \cdots b^i_1 \tilde{b}^i_0$
s^i_2	$(p_i - 1) mod\, 2^{n_i}$	$b^i_{n_i-1} \cdots \tilde{b}^i_1 b^i_0$
$\vdots$	$\vdots$	$\vdots$
s^i_k	$(p_i - / + a) mod\, 2^{n_i}$	$b^i_{n_i-1} \cdots \tilde{b}^i_{k-1} \cdots b^i_1 b^i_0$
$\vdots$	$\vdots$	$\vdots$
$s^i_{2^{n_i}-1}$	$(p_i - 2^{n_i-1}) mod\, 2^{n_i}$	$\tilde{b}^i_{n_i-1} \cdots \tilde{b}^i_1 \tilde{b}^i_0$

Generally speaking, LSB substitution belongs to BS strategy in which numbers of states pairs are specified. If s^i_0 and s^i_1 states in BS strategy are chosen for embedding, it is actually a one-bit LSB embedding. That is, at least 2 states in one parameter must be utilized for feasible steganography. When q states are selected, the obtained embedding capacity are $log_2 q$ bits. All the states depend on the current state of p_i and the performed operations of AS or BS on it. They stand for modified parameter values of p_i caused by steganography, which will influence the speech quality. And the essential of state-based embedding procedure is a map between current state and a set of available states relative to p_i with minimal speech quality distortion.

2.3 States Evaluation of G.723.1 Frame

In LSB substitution, every bit needs to be evaluated in order to get its noise (speech quality distortion) resistant characteristic. That is often performed by bitwise flipping on speech frames. Differing from LSB based method the state-based steganography is fine-grained and more accurate. The least significant states are evaluated state-wise according to current parameter states. AS strategy is an arithmetic based states method in which all states are obtained through arithmetical operations of the parameter. And states in BS strategy base on fine-grained bit modifications relative to current parameter values.

LSB embedding is essentially a BS strategy using 2 states as a unit. Therefore, as aforementioned, the embedding procedure is a transformation from original state to another state. The translated states relative to original one with least distortions are evaluated in Figure 2. Here the distortions are evaluated by commonly used perceptual evaluation speech quality (PESQ) method [8] which is proposed by ITU-T to objectively evaluate speech quality with a value between -0.5 and 4.5. Higher value means better speech quality or less distortion for steganography. The PESQ value threshold is set to 3.3 in the figure, which sounds good for human auditory system. Since only the several least significant bits are typically used for steganography and for brevity, we consider the first 11 states (S0...S10) at maximum in AS and BS strategies for our evaluation, and we present the result sitting above the threshold. In the figure, S0 stands for unmodified state which is equal to the PESQ value without steganography. Numbers in parentheses immediately after the parameters represent the numbers of selected states.



(a) Best states of AS strategy

(b) Best states of BS strategy

Figure 2: States test of G.723.1 speech frame

2.4 States Coding for Embedding and Extracting

The selected states need to be coded for applicable embedding and extracting. As the minimum processing unit of commonly used information is one bit, at least two states ($log_2 2$ bit) in one parameter are necessary. And at most 75 states in AS and 86 states in BS are available for hiding according to Figure 2. Let $g(x)$ denote the number of elements in set x. S_s represents a selected states set for steganography and S_{p_i} is a set consisting of all the states of p_i. $S_{s\cap p_i} = S_s \cap S_{p_i}$, $h_s = g(S_s)$ and $h_{s_i} = g(S_{s\cap p_i})$. The following conditions are necessary.

Condition 1 $2 \leq h_s \leq \sum_{i=0}^{23} 2^{n_i}$. (Here, maximal h_s are set to 75 and 86 for AS and BS, respectively.)

Condition 2 If $\exists\, s_j \in S_{s\cap p_i}$, then there must exist s_k satisfying $s_k \in S_{s\cap p_i}$, where $k \neq j$ and $s_j \not\equiv s_k \, mod \, h_{s_i}$.

Therefore, multiple multi-ary numeration systems can be utilized for embedding in different parameters. An h_{s_i}-ary digit is a number which contains h_{s_i} basic states corresponding to h_{s_i} states in $S_{s\cap p_i}$. Consequently, the embedding procedure in p_i (intra-parameter) falls into the following steps.

Step 1 Translate secret message M into an n digits h_{s_i}-ary number $(M)_{h_{s_i}} = \{m_{n-1} \cdots m_k \cdots m_1 m_0\}$, where m_k is an h_{s_i}-ary digit of $(M)_{h_{s_i}}$ and $m_k \in H_{s_i} = \{0, 1, \ldots, h_{s_i} - 1\}$.

Step 2 Build a map $E : H_{s_i} \rightarrow S_{s\cap p_i}$ which belongs to a map set with $A_{h_{s_i}}^{h_{s_i}}$ (full permutation) elements. Then m_k is mapped to s_k, where $s_k \in S_{s\cap p_i}$.

Step 3 Perform the modification related to s_k in AS or BS according to current states of p_i and E.

Otherwise, one single digit in M can be also embedded in different parameters (inter-parameter), that is the map $E' : H_s \rightarrow S_s$. The two schemes have the same embedding capacity. Assume $h_k, h_l \in H_s, k \neq l$. Then h_k and h_l may be mapped to 2 states s_k and s_l, where $s_k \in S_{s\cap p_k}$ and $s_l \in S_{s\cap p_l}$. And the embedding and extracting procedures are similar to intra-parameter scheme. Let p'_i represents a parameter with secret information (stego parameter). The extracting procedure for intra-parameter scheme is described as follows.

Step 1 $h'_{s_i} = g(S_{s\cap p'_i})$, if $h'_{s_i} \neq 0$, translate p'_i into an h'_{s_i}-ary number $(p'_i)_{h'_{s_i}}$.

Step 2 Get secret message digit m_j according to the reverse map of E and least significant h'_{s_i}-ary digit of $(p'_i)_{h'_{s_i}}$.

Step 3 Translate m_j into binary digits which are the secret message bits.

3. PERFORMANCE ANALYSIS

In both AS and BS strategies, intra-parameter and inter-parameter schemes need to employ different multi-ary embedding strategies with regard to parameters containing different numbers of selected states. And the two schemes bear the same embedding capacity for the embedding strategy. As each state is only relative to certain specific parameter, the states in one parameter are independent from others. Thus embedding capacity are deduced as $E_c = \sum_i log_2 h_{s_i}$. Accordingly, the capacities of AS (E_c^{AS}) and BS (E_c^{BS}) are figured out as follows.

$$E_c^{AS} = 4log_2 2 + 4log_2 3 + 5log_2 11 \approx 27.6 \texttt{ bits}$$
$$E_c^{BS} = 5log_2 2 + 4log_2 3 + log_2 4 + log_2 5 + 5log_2 11 \approx 32.9 \texttt{ bits}$$

3.1 Perceptual Evaluation

In order to perceptually evaluate proposed AS and BS embedding strategies and compared them with LSB methods, we test the average PESQ values of them with embedding rate at 100% in Figure 3. Two kinds of LSB substitution methods with 28 and 31 bits (approximately equal to AS and BS) embedding capacities are adopted. In the figure CM, CW, EM, EW represent Chinese woman, Chinese man, English man and English woman speeches which consist of 200 different pieces with the same duration of 10 seconds and sample rate of 8000 Hz, 16 bit. It can be seen that AS strategy outperforms similar LSB embedding (28 bits) with still acceptable speech qualities. And BS strategy has a little degradation but with most security performance that will be mentioned later. The experimentations performed on Intel Pentium Dual-Core E5200 2.5GHz CPU with 2GB DDR2 400MHz SD RAM exhibit respectively about 12.4 μs and 12.0 μs of average delays per frame, which are negligible compared with the 37.5 ms algorithmic delay of G.723.1 speech frame [7].

3.2 Security Consideration

Similar to Kerchoff's law in cryptography, security of steganography should not depend on the confidentiality of

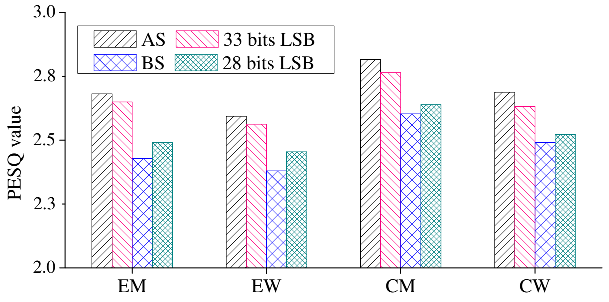


Figure 3: PESQ values comparison

coding algorithm but secure key. Compared with LSB based algorithm, state-based method employs more hiding possibilities (or capabilities), which increases the key space size either. The map set for embedding and extracting can be employed as the key space. In intra-parameter and inter-parameter strategies, the maximal key space sizes are respectively $(\prod_{\{h_{s_i}\}} A_{h_{s_i}}^{h_{s_i}}) \cdot A_{l_p}^{l_p}$ and $A_{h_s}^{h_s}$. Here l_p is the number of selected parameters in Figure 2, where $l_p = 13$ for AS and $l_p = 16$ for BS. And for LSB methods with approximately equal embedding capacity, the maximum key space size equals to $2^{E_c} \cdot A_{E_c}^{E_c}$. Table 2 lists maximum key space sizes of these methods. It is obvious that AS and BS strategies largely increase the key space size compared with corresponding LSB methods.

Table 2: Key space size of steganography

Algorithm	Embedding Scheme	Approx. Size
LSB	$E_c = 28$ bits	1.92×2^{125}
	$E_c = 33$ bits	1.58×2^{155}
AS	intra-parameter	1.09×2^{173}
	inter-parameter	1.32×2^{363}
BS	intra-parameter	1.96×2^{199}
	inter-parameter	1.87×2^{433}

4. CONCLUSION AND FUTURE WORK

This paper proposes a state-based steganography method including AS and BS strategies. Different from traditional bitwise embedding, the proposed method is fine-grained and takes one state ($log_2 1$ bit) as the minimal embedding unit. The coding and embedding procedures both base on the characteristics of states, which enhances embedding scalability and improves the performance. Future work will concentrate on the resistance of statistical steganalysis and the application of state-based method in other LSB based steganography algorithms.

5. ACKNOWLEDGMENTS

This work is supported in part by National Basic Research Program (973 Program) of China under Grant No.2011CB302305, National Key Technology R&D Program under Grant No.2012BAH35F03, Natural Science Foundation of Fujian Province of China under Grant No.2011J05151, and Scientific Research Foundation of National Huaqiao University under Grant No.11BS210.

6. REFERENCES

[1] W. FrÄĚczek, W. Mazurczyk, and K. Szczypiorski. Hiding information in a stream control transmission protocol. *Computer Communications*, 35(2):159–169, Jan. 2012.

[2] Y. Huang, S. Tang, and J. Yuan. Steganography in inactive frames of voip streams encoded by source codec. *IEEE Transactions on Information Forensics and Security*, 6(2):296–306, June 2011.

[3] J. Lubacz, W. Mazurczyk, and K. Szczypiorski. Vice over ip. *IEEE Spectrum*, 47(2):42–47, Feb. 2010.

[4] W. Mazurczyk and J. Lubacz. Lack-a voip steganographic method. *Telecommunication Systems*, 45(2):153–163, Oct. 2010.

[5] W. Mazurczyk, J. Lubacz, and K. Szczypiorski. Hiding data in voip. In *Proceedings of the 26th army science conference*, pages 1–4, 2008.

[6] W. Mazurczyk, P. Szaga, and K. Szczypiorski. Using transcoding for hidden communication in ip telephony. *Arxiv preprint*, abs/1111.1250, 2011.

[7] ITU-T Recommendation G.723.1. Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, May 2006.

[8] ITU-T Recommendation P.862. Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Feb. 2001.

[9] A. Shahbazi, A. H. Rezaie, and R. Shahbazi. Melpe coded speech hiding on enhanced full rate compressed domain. In *Proceedings of the 4th Asia International Conference on Mathematical Modelling and Computer Simulation*, pages 267–270, May 2010.

[10] H. Tian, K. Zhou, H. Jiang, and D. Feng. Digital logic based encoding strategies for steganography on voice-over-ip. In *Proceedings of the 2009 ACM Multimedia Conference, with Co-located Workshops and Symposiums*, pages 777–780, Oct. 2009.

[11] H. Tian, K. Zhou, and J. Lu. A voip-based covert communication scheme using compounded pseudorandom sequence. *International Journal of Advancements in Computing Technology*, 4(1):223–230, 2012.

[12] C. Wang and Q. Wu. Information hiding in real-time voip streams. In *Proceedings of the 9th IEEE International Symposium on Multimedia*, pages 255–262, Taichung, Taiwan, 2007.

[13] C. Wang, W. Zhang, J. Liu, and N. Yu. Fast matrix embedding by matrix extending. *IEEE Transactions on Information Forensics and Security*, 7(1):346–350, 2012.

[14] B. Xiao, Y. Huang, and S. Tang. An approach to information hiding in low bit-rate speech stream. In *Proceedings of 2008 IEEE Global Telecommunications Conference*, pages 1940–1944, New Orleans, LA, United states, 2008.

[15] H. Zhao, Y. Shi, and N. Ansari. Steganography in streaming multimedia over networks. *Transactions on Data Hiding and Multimedia Security VII*, 7110:96–114, 2012.