Digital Logic Based Encoding Strategies for Steganography on Voice-over-IP

Hui Tian[1], Ke Zhou[1]*, Hong Jiang[2], Dan Feng[1]
[1]Wuhan National Lab for Optoelectronics, School of Computer, Huazhong University of Science and Technology, Wuhan, 430074, China
[2]Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588-0150, USA
huitian@smail.hust.edu.cn, k.zhou@hust.edu.cn, jiang@cse.unl.edu, dfeng@hust.edu.cn

ABSTRACT
This paper presents three encoding strategies based on digital logic for steganography on Voice over IP (VoIP), which aim to enhance the embedding transparency. Differing from previous approaches, our strategies reduce the embedding distortion by improving the similarity between the cover and the covert message using digital logical transformations, instead of reducing the amount of the substitution bits. Therefore, by contrast, our strategies will improve the embedding transparency without sacrificing the embedding capacity. Of these three strategies, the first one adopts logical operations, the second one employs circular shifting operations, and the third one combines the operations of the first two. All of them are evaluated through comparing their prototype implementations with some existing methods in a prototypical covert communication system based on VoIP (called StegVoIP). The experimental results show that the proposed strategies can effectively enhance the embedding transparency while maintaining the maximum embedding capacity.

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
Algorithm, Design, Experimentation, Security

Keywords
Steganography, Digital Logic, VoIP, Transparency

1. INTRODUCTION
Steganography, a technology of information hiding, has been employed increasingly and successfully in covert exchange of secret messages [1, 2], watermarking and copyright protecting [3], and other practical applications. However, most of the previous studies on steganography are carried out on storage media (e.g. image, audio, etc) [4] and by contrast the area of steganography in streaming media is largely unexplored. However, as a cover of steganography, streaming media possesses two advantages over storage media. First, the real-time nature potentially provides better security for secret messages by virtue of its instantaneity, because it does not give eavesdroppers sufficient amount of time to detect possible abnormity due to hidden messages. Second, streaming media is a multidimensional cover, in which the packet protocol headers and the payload data can be used to hide data. Given the potential advantages, steganography for real-time streaming media may soon become a worthy subject of further studies. In this study, we will focus on a typical streaming media, Voice over IP (VoIP), as a possible carrier to apply steganography.

VoIP is a promising technique to enable telephone calls via a broadband Internet connection. Owing to its advantages of low cost and flexible advanced digital features, VoIP has become a popular alternative to public-switched telephone network (PSTN), and extensive research on it has been conducted [5]. In recent years, many researchers have carried out useful research on steganography over VoIP [6-12]. Most of them [6-8] mainly focus on the embedding methods or implementations of steganography on VoIP. Our practical experience and observation suggest that the encoding problem for VoIP-based steganography is also very important. Generally speaking, encoding strategies can be divided into four categories according to their objectives.

1. **Capability-Orientated Strategy (COS)**. COS aims to increase the information capacity of the covert message. It often involves certain compression codes, such as T-code, Run Length Code, etc.
2. **Reliability-Orientated Strategy (ROS)**. ROS aims to assure the correct extraction of covert messages at the receiver side. It often involves certain check codes [10] (e.g. Checksum, CRC, MD5, etc) or error-correcting codes (e.g. Hamming code, BCH code, Turbo code, etc).
3. **Security-Orientated Strategy (SOS)**. SOS aims to enhance the security of the covert message, namely, to avoid or minimize the possibility of unauthorized extraction. In [8], the authors introduced traditional cryptographies for embedded messages. However, traditional cryptographies are often time-consuming and incur delays that may in turn degrade speech quality drastically, so they do not suit this case well. In our previous works, we employed pseudorandom binary sequences, namely, the pseudorandom sequence produced by an improved Mersenne Twiste algorithm [10] and the m sequence [12], to encrypt the covert messages with the Exclusive OR operation (XOR), which can effectively balance between adequate security and low latency for real-time services.
4. **Transparency-Orientated Strategy (TOS)**. TOS aims to reduce the distortion and thereby enhance the transparency of steganography. Generally, a given steganography method is often believed to possess an inherent transparency. However, we argue that the transparency of the given steganography method can be improved by proper encoding. In this paper, we proposed three encoding strategies used to enhance the

MM'09, October 19 –24, 2009, Beijing, China.

transparency of the well-known Least-Significant-Bits (LSBs) steganography over VoIP.

Huang et al. [9] first introduced LSB matching method to enhance transparency. In their method, a pseudo random sequence (PRS) is used to guide embedding. If the current value of the PRS is "1", the corresponding LSB is replaced with one bit of secret messages; otherwise, the corresponding LSB is not modified. Due to the small amount of substitutions, this method can reduce the distortion of the cover speech in comparison with traditional LSBs steganography, but at the expense of halving the maximum capacity. Moreover, in [11], we introduced the notion of Partial Similarity Value (PSV) that is the similarity between LSBs and covert messages. By properly setting the threshold PSV, this approach can adaptively balance between the embedding transparency and the embedding capacity. Although this approach emphasizes on the tradeoff between transparency and capacity, it indicates that the embedding transparency can be enhanced by taking into account the similarity between the covers and the covert messages. Further, our observations through research and experiments suggest that the similarity between the covers and the embedded messages can be significantly increased by taking into account some transforms of the embedded messages. Therefore, we are motivated to propose the encoding strategies based on digital logic transforms for LSBs steganography over VoIP, which can reduce the distortion of the embedding and enhance the embedding transparency as a consequence while maintaining the maximum embedding capacity.

The rest of this paper is organized as follows. Section 2 presents the encoding strategies based on digital logic. Section 3 describes prototype implementations of the proposed strategies and some test results, and Section 4 gives concluding remarks and future works.

2. DIGITAL LOGIC BASED ENCODING

As mentioned above, we want to increase the similarity between the cover and the covert message by virtue of a given set of transforms, denoted by $F = \{f_1, f_2, ..., f_r\}$, where r is the number of the transforms. Let us assume that the covert message M is to be embedded into the cover C. In order to obtain the best similarity, we perform all transforms in F on M and get the set $F(M) = \{f_1(M), f_2(M), ..., f_r(M)\}$. Then, we evaluate the similarity between C and each element in $F(M)$, and choose the optimal element $f(M)$ with the best similarity. Accordingly, we can embed $f(M)$ into C to obtain the best transparency. Obviously, we should extract M by performing the inverse transform of transform f at the receiver side, which suggests that each adopted transform $f_i \in F$ must have the inverse transform f_i^{-1} and consequently $f_i^{-1}(f_i(M)) = M$. In addition, the computing complexities of the transforms must be small enough to meet the requirement of real-time services, which is why we employ the digital logic as the encoding transforms. In a real-time application, we often adopt "divide and rule" strategy, because M is often too large to be considered as a group. In other words, we divide C and M into N parts respectively, i.e. $C = \{C'_1, C'_2, \ldots, C'_N\}$ and $M = \{M'_1, M'_2, \ldots, M'_N\}$, where $C'_i = \{c'_{i1}, c'_{i2}, \ldots, c'_{il(i)}\}$, $M'_i = \{m'_{i1}, m'_{i2}, \ldots, m'_{il(i)}\}$, $l(i)$ is the length of the ith part, $i = 1, 2, \ldots, N$. Generally, the length of C is greater than or equal to the length of M. Here, we assume that their lengths are equal for the convenience of the description. Therefore, the lengths satisfy the following equation:

$$L_M = L_C = \sum_{i=1}^{N} l(i) \tag{1}$$

where, L_M and L_C are the lengths of M and C respectively. Accordingly, the transforms will be performed in parts and partial similarity value (PSV) [11] will be introduced to evaluate the similarity. For the LSBs steganography, PSV indicates the number of identical bits between the message part and the corresponding cover part. The following text will present our digital logic based encoding strategies in detail.

2.1 Encoding Based on Logical Operations

Table 1. The transforms definition of strategy 1

No.	Op.	Transform	No	Op.	Transform
1	ORI	ORI(M'_i)	**4**	NOT	NOT(S'_i)
2	ORI	ORI(S'_i)	**5**	XOR	XOR(M'_i, S'_i)
3	NOT	NOT(M'_i)	—	——	——
Note: (1) $M'_i \in M$ and $S'_i \in S$; (2) ORI(x) = x, $x = M'_i$ or S'_i.					

As is well known, the common logical operations include AND, OR, XOR and NOT. However, we can only choose XOR and NOT, because the others have no inverse operations. Moreover, since it is performed between two elements, we need to introduce a PRS as the reference sequence (RS) when using XOR. Typical PRS candidates include m sequence [12, 13], chaotic sequence [14], sequence generated by toral automorphisms [15], etc. We denote RS as $S = \{S'_1, S'_2, \ldots, S'_N\}$ and still assume that the length of S (denoted by L_S) is equal to the length of M and C, i.e. $L_S = L_M = L_C$. We can define the transforms in Table 1.

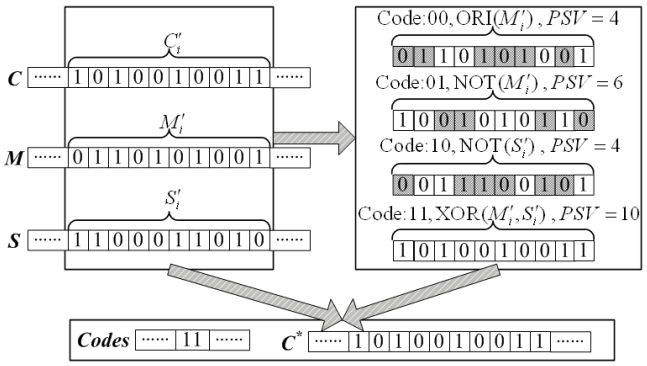

Figure1 The embedding process using strategy 1

As mentioned above, we perform each transform on the given part, and choose its optimal form to embed into the cover. In order to inform the adopted transform to the receiver, we set a binary code for each transform. The length of each code (denoted by L_{code}) can be determined by the following formula:

$$L_{code} = \lceil \log_2(Q^*) \rceil \tag{2}$$

where, Q^* is the number of the adopted transforms. If we choose all the transforms in Table 1, then $L_{code} = 3$, which means there are 3 ($2^3 - 5$) binary codes left to be unused. Therefore, we prefer to choose four transforms among them to get the optimal $L_{code} = 2$. Figure 1 illustrates the embedding process, in which transforms 1, 3, 4 and 5 are adopted. Another problem is how to transmit these indicating codes to the receiver. In our previous study [12], we presented a synchronization mechanism using techniques of the protocol steganography, in which synchronization patterns are hidden in the unused and/or optional fields of the header of a certain packet. In the IP header, there are a total of 64 bits that can be used to embed messages [16]. Moreover, the headers of upper-level protocols (e.g. RTP, etc) also have many unused or optional fields. Therefore, we can distribute the codes among those fields in a predetermined manner. We will give a specific example in

Section 3 to illustrate this approach. The receiver knows exactly how and where the codes are embedded. The receiver first checks the codes and further extracts the corresponding hidden parts when receiving an IP packet. After collecting all the parts, the receiver can successfully restitute the whole secret message.

2.2 Encoding Based on Shifting Operations

The shifting operations, which allow bits to be moved to the left or right in a word, are often used in serial transfer of data. There are three types of shifting operations: logical, arithmetic and circular. The first two types cannot be adopted in our strategy, for they have no inverse operations that can be used to get the original data. However, as its name implies, the circular shift circulates the bits in the given word around the two ends without loss of information, so we can recover the original data by the opposite shift. The circular shift involves the circular shift left (CSL) and the circular shift right (CSR). For an x-bit word, a CSL by x-y bits (denoted by CSL(x-y), $0 \le y \le x$) is equivalent to a CSR by y bits (denoted by CSL(y)). Therefore, we define the circular shift based transforms that consist of the operation codes (Op. code) and the shift numbers. Figure 2 depicts a set of transforms for 8-bit words. As the figure shows, the first 1 bit marked with underline in each code is used to represent the Op. code, namely Op. = 1 represents CSR; and Op. = 0 represents CSL. In this strategy, the length of each code (L_{code}) depends on the operation word (one part of the message) and can be determined as follows:

$$L_{code} = \lceil \log_2(n) \rceil \tag{3}$$

where, n is the length of the given part. The rest process can be performed similarly as in strategy 1.

Code	Description	Code	Description	Code	Description	Code	Description
000	CSL (0)	001	CSL (1)	010	CSL (2)	011	CSL (3)
100	CSR (4)	101	CSR (3)	110	CSR (2)	111	CSR (1)

Figure 2 The circular shift based transforms for 8-bit words

2.3 Hybrid Strategy

We can merge the two strategies above. A straightforward method may be to merge these transforms directly. However, due to the difference between these transforms' characteristics, the codes of unequal length have to be adopted. Namely, the codes for the transforms in strategy 1 are tantamount to the Op. codes, and the codes for the transforms in strategy 2 include the Op. codes and the shift numbers. This will incur more cost in bits to represent the code and induce an intractable synchronization problem. Thus, the value of this method is limited. Another method is to first apply the two strategies in each packet and choose the one with the larger total PSV as the final strategy. This method needs extra 1-bit flag to indicate which strategy is finally adopted. However, due to the option of more transforms, we can further enhance the transparency for the embedding process.

3. IMPLEMENTATION AND TEST

This section will give specific examples illustrating the proposed strategies and evaluate them in StegVoIP that is a prototypical covert communication system based on VoIP [10-12]. StegVoIP supports typical coders, such as G.711, G.723.1, G.729a, etc. In the prototype implementations, we typically adopt G.729a as the codec of the cover speech, while the strategies can also effectively work in conjunction with other coders. Similarly, we choose 8 LSBs (the bits with the least impact on the speech quality) in each G.729a frame based on the observation that the parameters of fixed codebook have the best transparency for data hiding [11, 12]. We consider the embedding process in each packet. Let the length of codes in each packet be X, the number of LSBs be Y. If the Y LSBs is divided into n parts and the length of the ith part is l_i, for our strategy 1, we can obtain the following equations:

$$\begin{cases} 2 \cdot n = X \\ \sum_{i=1}^{n} l_i = Y \end{cases} \tag{4}$$

And, we can obtain the following equations for our strategy 2:

$$\begin{cases} \sum_{i=1}^{n} \log_2 l_i = X \\ \sum_{i=1}^{n} l_i = Y \end{cases} \tag{5}$$

We employ the identification field of the IPv4 header to hide the codes of transforms. As proposed in [17], the high 8 bits of the identification field can be used for data hiding, so we can set $X = 8$. Moreover, we set the payload of each packet at 4 frames (40 octets) in order to not induce the fragment strategy of Internet protocol. Consequently, 32 bits are available for data hiding in each packet ($Y = 32$). There are many partition methods. In this paper, for strategy 1, LSBs are divided into 4 parts that contain 8 bits each and each code can be represented by 2 bits; for strategy 2, LSBs are divided into 2 parts that contain 16 bits each and each code can be represented by 4 bits. For strategy 3, we employ the second bit (DF) in the flag field, which represents "do not fragment" [16] [17]. Because the packets are not larger than the maximum fragment size, the setting of the DF flag has no effect on the packets' behavior. Here, DF = 0 indicates that strategy 1 is adopted, and DF = 1 indicates that strategy 2 is adopted.

We compare our strategies with the traditional LSBs method and the LSB matching method in [9]. We employ m sequences as the PRSs used in our strategy 1 and the LSB matching method. All m sequences are randomly generated. To compare the transparency of these methods, we define the bit-change rate (*BCR*) as follows:

$$BCR = \frac{N_C}{N_M} \tag{6}$$

where, N_C is the number of the changed bits, N_M is the number of the embedded bits. Particularly, *BCR* for the proposed strategies can be further described as follows:

$$BCR = \frac{\sum_{i=1}^{N} (l_i - \eta_i)}{N_M} \tag{7}$$

where, l_i and η_i are respectively the part size and the PSV of the ith replaced parts; N is the number of the replaced parts. Further, we employ the Perceptual Evaluation of Speech Quality (PESQ) method [18] to evaluate the speech quality of cover audios and their steganographic versions. PESQ compares an original signal with a degraded signal and outputs a PESQ score as a prediction of the perceived quality. The range of the PESQ score is -0.5 to 4.5. Moreover, the PESQ score can be converted to Mean Opinion Score – Listening Quality Objective (MOS-LQO). The range of MOS-LQO is 1.017 (worst) to 4.549 (best), which more closely matches the range of subjective Mean Opinion Score (MOS) [19]. For the experiments, we collect 320 ten-second speech samples [1]. These samples consist of two categories: English speech (including male speech and female speech) and

[1] Although ITU-T P.862 recommends that the length range of each test speech sample is 8 to 30s, PESQ is validated in ITU-T for use with signals that are mostly 8-12s long. Therefore, we typically choose the 10s long audio samples. In addition, so far, PESQ has not been validated with music as input to a codec, so our test samples do not include music.

Chinese speech (including male speech and female speech). All samples are PCM coded files with 8000 HZ sampling rate, 16 bits quantization and mono, which are the reference signals in the PESQ experiments. For each sample, the corresponding steganography experiments are performed on its G. 729a coded file respectively. The secret message (choosing the CALL FOR PAPERS file of ACM Multimedia 2009 [20], possibly only some forward parts) can be successfully embedded and retrieved in any case. We convert all the steganographic G. 729a encoded files into PCM encoded files as the degraded signals and perform the PESQ test. In the five methods, the embedding capacities of the traditional LSB and the proposed strategies are 8000 bits, but the average embedding capacity of the LSB matching method is 4000 bits. Figure 3 and Figure 4 show the statistical results of the mean *BCR* and the mean MOS-LQO for all the five methods.

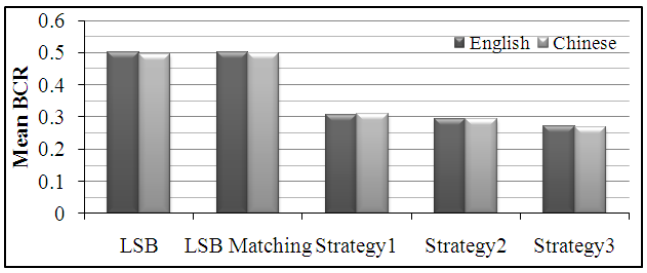

Figure 3 Test results of bit-change rate

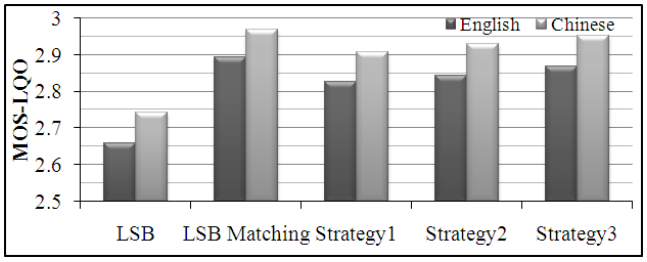

Figure 4 Test results of MOS-LQO

From these charts, we can observe that: (1) For the same method, the mean MOS-LQO of Chinese speech samples is slightly larger than that of English speech samples, which implies that Chinese speech samples may have better embedding transparency than English speech samples. (2) The LSB matching method has the largest MOS-LQO, but it halves the embedding capacity. By contrast, the proposed strategies in this paper reduce the distortion by decreasing the number of bits changed and achieve comparable MOS-LQO scores with that of the LSB matching while maintain the maximum embedding capacity.

In contrast with the traditional LSBs method and the LSB matching method, our strategies may increase the complexity of steganography. However, the proposed strategies will not impact the real-time services of VoIP, because only a few simple digital logic transforms are employed. The readers may find that the embedding transparency can be further enhanced if we introduce more transforms. However, more transforms translate into higher complexity of the algorithm and more indicating codes. Thus, a sensible tradeoff among them should be sought.

4. CONCLUSION AND FUTURE WORK

This paper presented transparency-orientated encoding strategies for steganography over VoIP. Differing from previous approaches, the proposed strategies employ digital logic operations to increase the similarity between the cover and the covert message, and thereby effectively reduce the distortion (enhance the transparency) while maintaining the maximum embedding capacity. From the experiments, we learn that the embedding transparency can be enhanced further if we introduce more digital logic transforms. However, more transforms will likely induce higher complexity of the algorithm and more indicating codes. Therefore, one must often strike a sensible tradeoff among the embedding transparency, the number of required indicating codes and the complexity of the algorithm, which will be one of our future works. In addition, we will study other encoding strategies, which aim at improving the embedding performance of steganography over VoIP.

5. ACKNOWLEDGMENTS

The work is supported partially by National Basic Research 973 Program of China under Grant No. 2004CB318201, Program for New Century Excellent Talents in University under Grant No. NCET-06-0650, US National Science Foundation under Grant No. CCF-0621526, 863 project under Grant No. 2009AA01A402 and Program for Changjiang Scholars and Innovative Research Team in University under Grant No. IRT-0725.

6. REFERENCES

[1] N. Provos, P. Honeyman. Hide and seek: an introduction to steganography, *IEEE Security & Privacy Magazine*, Vol. 1, Issue 3, May-June 2003, pp. 32-44.

[2] K. Bailey, K. Curran. An evaluation of image based steganography methods. *Multimedia Tools and Applications*, Vol. 30, Issue 1, July 2006, pp. 55-88.

[3] E. T. Lin, A. M. Eskicioglu, etc. Advances in digital video content protection, *Proceedings of the IEEE: Special Issue on Advances in Video Coding and Delivery*, Vol.93, No.1, January 2005, pp.171-183.

[4] M. Shirali-Shahreza. "A new method for real-time steganography", *Proceedings of the 8th International Conference on Signal Processing*, Vol. 4, pp. 16-20, 2006.

[5] B. Goode. Voice over Internet protocol (VoIP), *Proceedings of the IEEE*, Vol. 90, Issue 9, Sept. 2002, pp. 1495-1517.

[6] C. Wang, Q. Wu. Information hiding in real-time VoIP streams, *Proceedings of the 9th IEEE International Symposium on Multimedia*, 10-12 Dec. 2007, pp. 255-262.

[7] J. Dittmann, D. Hesse and R. Hillert. Steganography and steganalysis in voice over IP scenarios: operational aspects and first experiences with a new steganalysis tool set, *Proceedings of SPIE, Vol. 5681, Security, Steganography, and Watermarking of Multimedia Contents VII*, March 2005, pp. 607-618.

[8] C. Kratzer, J. Dittmann, T. Vogel, and R. Hillert. Design and evaluation of steganography for voice-over-IP, *Proceedings of 2006 IEEE International Symposium on Circuits and Systems*, 21-24 May 2006, pp. 2397-2340.

[9] Y. Huang, B. Xiao, H. Xiao. Implementation of Covert Communication Based on Steganography. In *Proceedings of 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Aug. 2008, pp. 1512 - 1515.

[10] H. Tian, K. Zhou, Y. Huang, etc. A Covert Communication Model Based on Least Significant Bits Steganography in Voice over IP. In *Proceedings of the 9th International Conference for Young Computer Scientists,* Nov. 2008, pp.647-652.

[11] H. Tian, K. Zhou, H. Jiang, etc. An Adaptive Steganography Scheme for Voice over IP. In *Proceedings of the 2009 IEEE International Symposium on Circuits and Systems*, Taipei, Taiwan, May 24-27, 2009, pp. 2922-2925.

[12] H. Tian, K. Zhou, H. Jiang, etc. An M-Sequence Based Steganography Model for Voice over IP. In *Proceedings of the 2009 IEEE International Conference on Communications*, Dresden, Germany, June 14-18, 2009.

[13] S. Engelberg, H. Benjamin. Pseudorandom sequences and the measurement of the frequency response. *IEEE Instru-mentation & Measurement Magazine,* Vol. 8, Issue 1, Mar. 2005, pp. 54 -59.

[14] A. Tefas, A. Nikolaidis, N. Nikolaidis, etc. Statistical Analysis of Markov Chaotic Sequences for Watermarking Applications. In *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems*, May 6-9 ,2001 pp.57-60

[15] G. Voyatzis, I. Pitas. Applications of toral automorphisms in image watermarking. In *Proceedings of the IEEE 1996 International Conference on Image Processing*, Sept.16-19, 1996, pp. 237 - 240

[16] S. J. Murdoch, S. Lewis. Embedding Covert Channels into TCP/IP. *Proceedings of the 7th Information Hiding workshop*, June, 2005, pp. 247-262.

[17] K. Ahsan and D. Kundur. Practical data hiding in TCP/IP. *Proceedings of the Workshop on Multimedia Security at ACM Multimedia*, Dec, 2002. pp. 63-70.

[18] 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.

[19] ITU-T Recommendation P. 800. Methods for subjective determination of transmission quality, Aug. 1996.

[20] Available at: http://www.acmmm09.org/ACMMM09_CFP.pdf.