GENERALIZED TRELLIS CODED QUANTIZATION FOR DATA HIDING

Xiaofeng Wang, Xiao-Ping Zhang

Ryerson University, Department of Electrical and Computer Engineering
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3
{xwang,xzhang}@ee.ryerson.ca

ABSTRACT

Information theory guides us to investigate the choice of quantizers for data hiding applications. In this paper, the design of the quantizer selection rule in trellis coded quantization (TCQ) based data hiding is discussed. A novel trellis branch quantizer selection rule which changes the old state transition function and takes advantage of all trellis states is proposed so as to increase robustness against attacks. Theoretical analysis and simulation results show that the new TCQ path selection (or branch selection) method achieves better bit error rate (BER) performance in the case of Gaussian attack compared to other popular approaches. The path selection rule could also be used as a secret key to provide security for the practical data hiding.

Index Terms— trellis coded quantization, data hiding, bit error rate

1. INTRODUCTION

In the past several years we have witnessed a big surge of the data hiding research somewhat due to the development of Internet. Quantization based embedding [1] and spread-spectrum [2] are two most commonly used data hiding methods. The simple quantization based method has the advantage of host interference rejecting and is a down-to-earth implementation of Costa's dirty paper theory [3]. A long-standing problem is the design of practical quantizers that is robust to attacks and introduces less distortion given certain amount of data to be embedded. Trellis coded quantization (TCQ) is a special case of vector quantization proposed by Marcellin and Fischer [4] which could achieve performance near rate distortion bound. Its structured code book provides more robustness to channel distortion. So it is suitable for data hiding purpose. Chou et al. [5] treat data hiding as channel coding with side information only available at the transmitter and employ TCQ to approach the theory bound. The authors in [6] explore the redundancy in initial state selection during trellis coded quantization to hide information and compare the result with quantization index modulation (QIM) and TCQ by path selection [5]. But the embedding rate of initial state selection is low. Our focus is the high data rate data hiding problem with independent identical distributed (IID) host.

This paper proposes a novel secure data hiding method based on TCQ by path selection. We analyze the TCQ based data hiding and show TCQ trellis can be designed to achieve more robust by changing the state transition rule and quantizer selection rule. And the new method gains the robustness benefit without increasing the complexity. We find that the rule of branch quantizers selection could be used as a key to provide the security to the data hiding with nearly the same distortion and robustness. Experiment results show that at the same distortion-to-noise ratio (DNR), the bit error rate (BER) for the new TCQ embedding method can be smaller than both the traditional TCQ embedding and QIM under Gaussian noise attack.

2. DATA HIDING USING TRELLIS CODED QUANTIZATION

2.1. Quantization based data hiding

In quantization based data hiding, the cover x with length n is represented by a quantized version u as shown in Fig. 1. There are several alternative quantized sequence subsets so the information sequence m could be embedded by choosing a specific sequence. The encoding is the process of the quantizer selection. For simplicity, we consider the transmission of n dimension binary sequence $m \in \{0,1\}^n$ and $\mathcal{U}$ is the set of all possible sequences. The communication process is as follows:

1. Identify all possible sequences $\mathcal{U}_m \subset \mathcal{U}$ that could represent message m.
2. Compute the closest sequence $u \in \mathcal{U}_m$ to x.
3. The quantized sequence is corrupted by attacks. For example additive Gaussian noise v with mean 0 and variance δ_v^2, the channel output is $y = u + v$.
4. Detect the hidden message $\hat{m}$ from y by searching most possible subset $\mathcal{U}_{\hat{m}}$ that y may belong to.

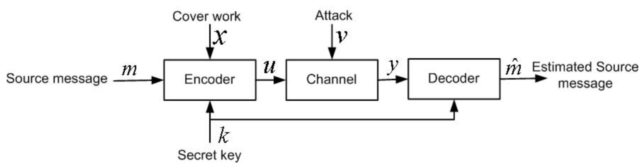


Fig. 1. Basic data hiding model.

One practical implementation QIM has gained popularity since it is proposed by Chen and Wornell [1]. In QIM, if the quantization step is set to $\Delta, \mathcal{U}_m$ is defined by

$$\mathcal{U}_m = \mathcal{U}_m[1] \times \mathcal{U}_m[2] \times \ldots \mathcal{U}_m[n], \tag{1}$$

$$\mathcal{U}_m[i] = \{k\Delta + \frac{\Delta * m[i]}{2},\ k \in \mathbf{Z}\}. \tag{2}$$

2.2. Trellis coded quantization

The TCQ[4] as shown in Fig. 2 is a quantization technique which borrows set partition idea from TCM (trellis coded modulation) to have better rate distortion performance in compression. In this figure each branch of trellis is associated with one quantizer. For example the first branch of state 0 is D_0. There are altogether 4 quantizers D_0, D_1, D_2 and D_3 which are divided into two groups $A_0 = D_0 \cap D_2$ and $A_1 = D_1 \cap D_3$. At each state, we could choose one of two groups of quantizers to quantize the source input x. In TCQ we pick up connected branches (quantizers) in trellis by using Viterbi algorithm to quantize a sequence so as to have less accumulated distortion.

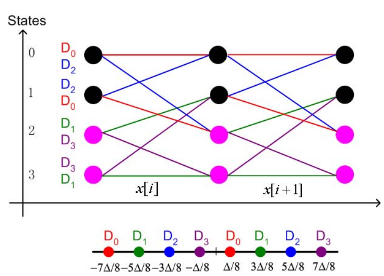


Fig. 2. Two steps of four states trellis coded quantization.

2.3. TCQ for data hiding

Since TCQ could provide some kind of robustness to the compression, it could be considered one of best candidates for data hiding. We treat QIM as a simple TCQ data hiding method without sate selection. It has only one state and two branches (two quantizers) as shown in Fig. 3. The set partition in QIM is regular. In TCQ data hiding, the quantizer selection is a function of states,

$$\mathcal{U}_m[i] = \{k\Delta + g(m[i], s_i),\ k \in \mathbf{Z}\} \tag{3}$$

$$s_i = t(s_{i-1}, m[i-1]) \tag{4}$$

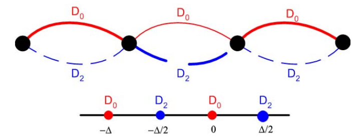


Fig. 3. An example of QIM embedding [0 1 0].

where s_i is the current state which is a function t of former states s_{i-1} and former embedding bit $m[i-1]$ and g is the branch selection rule which defines the quantizer to be used. So it creates a pseudo-random space partition which is preferable against attacks. In commonly used method [5][6] the first branch of each state is used to transmit data bit 0 and the second branch for embedding bit 1 as shown in Fig. 4(a).

3. NEW TRELLIS PATH SELECTION RULE FOR DATA HIDING

Quantization based data hiding could be explained in TCQ frame work. The embedding process is to find the minimum distortion path through the trellis given the data sequence m,

$$u = \arg\min_{u \in \mathcal{U}_m} \|u - x\|^2. \tag{5}$$

The detection process is to search minimum distortion path to the attacked sequence y,

$$\hat{u} = \arg\min_{\hat{u} \in \mathcal{U}} \|\hat{u} - y\|^2. \tag{6}$$

If $\hat{u} \in \mathcal{U}_{\hat{m}}$, then we extract the embedded data $\hat{m}$. There are two basic components in the TCQ data hiding, the path selection function $g(\cdot)$ and the state transition function $t(\cdot)$. The problem is how to design these two functions so as to have better tradeoff between robustness and distortion given the data payload. The data hiding introduced distortion is determined by the embedding processing. The common path selection method employ the same trellis structure as TCQ and use the input of the state machine to decide the embedding bit. It does not take full advantage of the trellis branch selection. For example in Fig. 4(b) the data sequence [1 0 1] is transmitted by using trellis as Fig. 4(a). In step 1 we embed 1, using the state transition $t(s_{i-1}, 1)$ s_i will be 2 or 3. In step 2 we hide 0, the s_{i+1} could only be 1. In step 3, s_{i+2} will be 2. So after two steps there is only one path available.

Because TCQ data hiding depends on states and the data to select quantizers, if the common method is used the branch selection is limited to certain states. There is no alternative path available. It means it lose the property of TCQ which reduces the distortion by delaying the choosing of quantizers. Actually it is not necessary that the data hiding is done by using the commonly used trellis state transition as in TCQ and TCM. When we do compression we need the input of the state transition machine to decide the next state. But in data

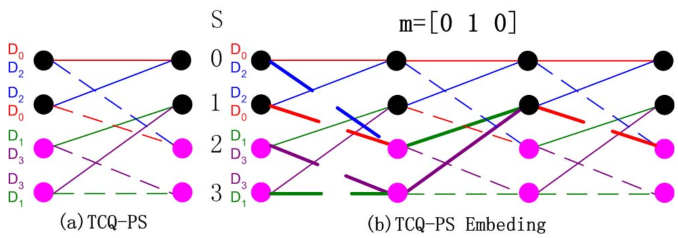


Fig. 4. An example of using traditional TCQ path selection method to hide data [0 1 0].

hiding applications the distortion and the robustness are our main concerns. The compression rate of the source has no direct relationship with our goals. So the trellis could be redesigned to incorporate more advanced trellis configuration. We show two new configurations to embed data into 4 states trellis in Fig. 5(a)(b). The solid line indicates embedding 0 and the dash line means quantizers for 1. The trellis structure is the same as the commonly used one. But the path selection and state transition rule are specially designed. The TCQ-PS-NEW in Fig. 5(a) is the best because it could use all 4 states of the trellis and it increases robustness by separating the two groups of quantizers further. The computational complexity is not increased for the new configuration because the basic trellis structures are kept all the same.

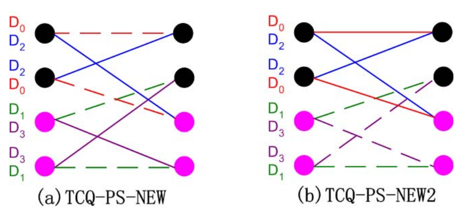


Fig. 5. New TCQ based data hiding method.

By choosing a new path selection function, we could have less distortion without losing robustness property. So the path selection function could be used as a key to increase the security of data hiding.

4. GENERALIZED TCQ DATA HIDING

We design a new trellis coded quantization data hiding system with path selection rule as secret key shown in Fig. 6. The embedding procedure is as follows.

- Step 1 - Initialization: set the overall distortion measure for all states to 0 and select the quantizers and the trellis path selection rule as the secret key.
- Step 2 - Quantization: if the input bit $m[i]$ is 0, the source $x[i]$ is quantized by 0 trellis quantizers (trellis with only the solid line). Otherwise the dash line is used.
- Step 3 - Branch selection: The distortion of each state is measured by selecting the minimum accumulated distortion branch. The distortion is the branch quantization distortion adding the former state's distortion.
- Step 4 - Repeat step 2 and 3 until the design trellis length is reached. The minimum distortion path is selected as the output.

The trellis branch selection function and state transition function could be randomly selected as long as the minimum path could be discriminated from other paths. At the detector side, the information is extracted by using the Viterbi decoding through the whole trellis.

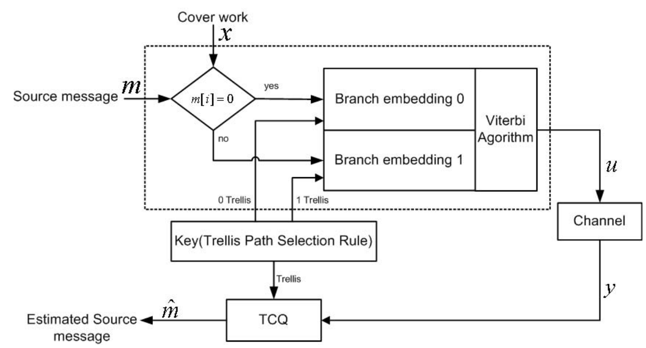
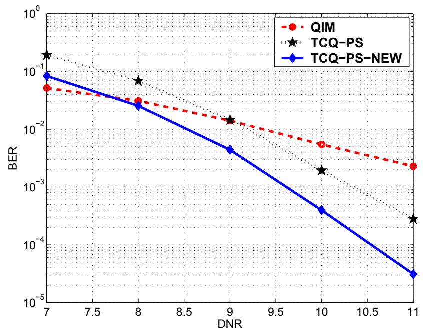


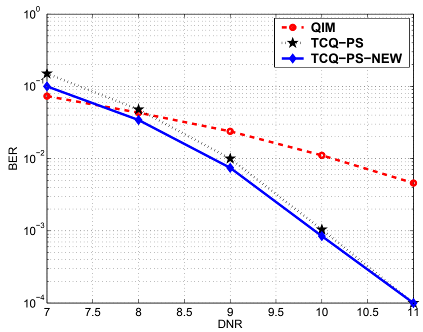
Fig. 6. Structure of TCQ embedding and detecting.

5. SIMULATION RESULTS

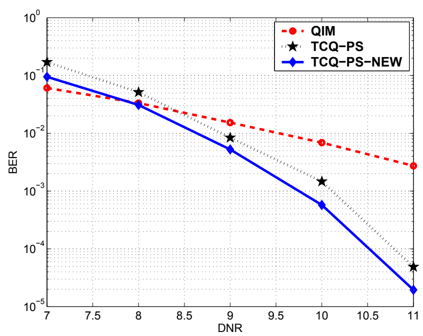
We apply the new trellis path selection rule of 4 states to different kind of sources with gaussian noise attack. In the Fig. 7 shown below, the source is laplacian distribution, gaussian distribution and the real image wavelet coefficients and the data to be embedded is 0 and 1 with the same probability. The embedding rate for all cases is the same 1 bit per cover-element. We compare the bit error rate performance of new TCQ method TCQ-PS-NEW as Fig. 5(a) with the traditional method TCQ-PS as Fig. 4(a) and QIM. The DNR is defined as the of the distortion to noise ration. The distortion is the quantization introduced distortion and the noise is added accordingly. The BERs are computed by averaging the bit error rate of 1000 random experiments. The trellis length in laplacian and gaussian case is 256 and in the image case the trellis length of 1024 is used. In low DNR QIM is the best. But increasing DNR to 8 db and over all TCQ cases are better than QIM. The simulation shows that the TCQ-PS-New new path selection rule is the best in term of BER performance. It is about 0.8db better than the common TCQ method in Laplacian case at BER 10^{-3}. For image wavelet coefficients it is only 0.4db better. It is because the trellis is longer though image coefficients are laplacian distributed. In gaussian source case it is 0.2db better at BER 10^{-2}.



(a)Laplacian source with mean 0 and variance 1



(b)Gaussian source with mean 0 and variance 1



(c)Image data hiding in the wavelet domain of Hill image

Fig. 7. Comparison among different TCQ path selection rules and QIM.

We apply the generalized TCQ method to images in wavelet domain by only selecting the 3nd diagonal subband coefficients for embedding. The coefficients are sent into diagram shown in Fig. 6. Fig. 8 shows typical images (gray level Hill and Lena of size 256x256) after the wavelet reconstruction with embedding PSNR of 48db and 43db and totally 1024bits are embedded. For (a)(b) the trellis TCQ-PS-NEW in Fig. 5 is used. For (c) TCQ-PS-NEW2 in Fig. 5 is used. There is no visual difference between (a) and (c). But if the key is unknown to the detector, the BER will be around 50%. In Fig. 8(d) we visualize the error pattern of using TCQ-PS-NEW to extract data from image Fig. 8(c). It shows that the rule could be used as a key to improve the security of data hiding.

Fig. 8. Example images of the generalized TCQ embedding and the error pattern of using the wrong key.

6. CONCLUSION

A new trellis coded quantization method is developed for data hiding. By analyzing the data hiding problem we realize that branch selection rule could be specially designed to reduce distortion. Simulation results show that the new trellis configuration has better performance than the commonly used one and the path selection rule could be used as a key for practical use. The presented data hiding scheme is distinguished by its ability to achieve less distortion without affecting robustness and increasing the complexity. Future work may include investigating more sophistical trellis structure.

7. REFERENCES

[1] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May 2001.

[2] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[3] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[4] M.W. Marcellin and T.R.Fischer, "Trellis coded quantization of memoryless and gauss-markov sources," *IEEE Trans. Commun.*, vol. 38, pp. 82 – 93, Jan. 1990.

[5] J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between data hiding and distributed source coding," in *Proc. 33rd Annual Asilomar conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1999, vol. 2, pp. 1503–1507.

[6] E. Esen and A. A. Alatan, "Data hiding using trellis-coded quantization," in *IEEE ICIP*, Singapore, Oct. 2004, vol. 1, pp. 59–62.