Network Watermarking Location Method Based on Discrete Cosine Transform

Xueyan Hou[1,2], Yonghong Chen[1,3,*], Hui Tian[1,4], Tian Wang[1,2], Yiqiao Cai[1,3]

[1]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, Fujian, CHINA

[2]{hollyhxya,wsnman}@gmail.com, [3]{*djandcyh,yiqiao00}@163.com, [4]htian@hqu.edu.cn

Keywords: network watermarking; DCT; discrete watermark; location

Abstract. The idea of network digital watermarking comes from the audio digital watermarking, which plays an important role in traffic tracking. However, there are many problems in the existing watermarking schemes. For example, the selection of watermark embedding area is not combined very well with the characteristics of the data stream; the interaction between watermarks reduces the detection rate and the concealment. In this paper, we propose a method of using the Discrete Cosine Transform(DCT) domain to locate the watermarks. After the discrete cosine transform, we select the lower energy area for discrete embedding watermarks. This is a novel attempt for the DCT in network watermarking. The experimental results show that the proposed method DCTBWL (Discrete Cosine Transform Based Watermarking Locating) can effectively improve the concealment and robustness of the watermarks in network flow.

Introduction

With the rapidly development of internet, network security issues have become more and more seriously. Stepping stone[1], anonymous communication system and botnet[3] make it very hard to trace the attacker. In order to better solve the above problems, many scholars have put forward the active network flow watermark(ANFW) technology by drawing on the idea of digital watermarking[2]. Some special information will be hidden through changing some characteristics of the flow generated from the sender. After the network transmission, if the corresponding watermark is detected at the receiver, it is considered that there is an obvious network communication relationship between the sender and the receiver. This is a kind of active network flow shaping and analysis technology.

Nowadays the embedding of watermarks always have nothing to do with the content of the packets. According to the different watermark carrier, ANFW technology mainly includes traffic rate[4], packet timing[5] and interval centroid[6~8]. However, the existing methods are not perfect, there are some problems as following: 1), the selection of watermark embedding area is not very well combined with the characteristics of the data stream. For example, the offset in [6] and [7] is only considered the time aspect and the characteristic of the flow is neglected. 2), The watermarks are embedded in a continuous area, when the front of the stream appears to be inserted or lost, all of the watermarks behind will be affected. Which result in the decrease of the watermark detection rate([5~8]). 3), It is easy for attacker to recognize the abnormity when a continuous area of data has been changed. With the low concealment, watermarks are easy to be removed or damaged. Such as the exploration for ICBW and IBW[9,10], or the exploration for DSSS-W[11,12].

Based on the above problems exist in the current watermarking schemes, in this paper, we propose a new method for network watermark location by using the discrete cosine transform(DCT). The experimental results show that this method has more robust and better concealment.

The rest of this paper is organized as follows: Section 2 analyzes the feasibility of the application of DCT in network watermarking as well as the process in detail. The experimental results validating the analysis are presented in Section 4. The paper is concluded in Section 5 along with some future research directions.

Discrete Cosine Transform Based Watermarking Locating Scheme

Feasibility Analysis of the Application of DCT in Network Watermarking. Digital watermarking technology plays an important role in the field of copyright protection and information security. The discrete cosine transform is widely used in image watermarking and audio watermarking technology[13], and the distribution of energy is presented after transform. Since it is difficult to perceive the operation in the lower energy area; the information can be hidden with high concealment. The idea of network digital watermarking comes from the image and audio watermarking technology, which is similar to a certain extent. Extracting an feature of network flow as an input signal, and then we could have the distribution of its energy by DCT. Pick up the original flow which with low energy as the watermark embedding area. Experimental results show that the discrete watermark embedding model reduces the interaction between the watermarks, and improves the robustness of the watermark in the transmission process.

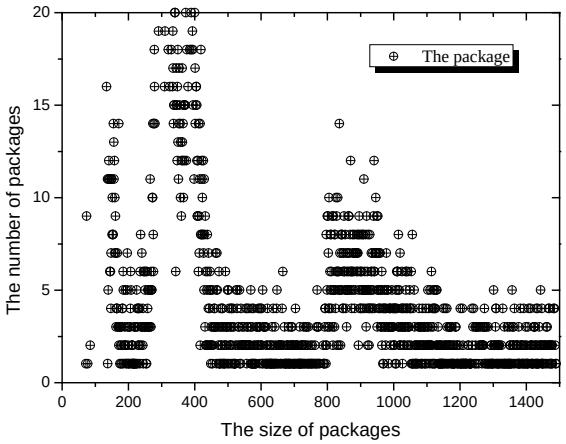

Fig. 1. The distribution of package size

In this paper, we used the data stream which sent from 68.142.235.31 to 50.21.229.205 in Chicago on March 20, 2014 as the research object. As the size of a package can reflect how much information contains directly, the size can be regarded as the characteristic of the package. Fig. 1 shows the distribution of package size from 1:06 pm. to 1:10 pm.. It is easy to find that the size is distributed over a large range. Different size of the package provides the possibility of energy distribution.

Discrete Cosine Transform Based Watermarking Locating Scheme. The discrete watermark embedding process consists of three parts: calculating the discrete cosine transform coefficient, selecting area and embedding watermark bits.

A. Discrete cosine transform

Given a packet flow of duration $T_f > 0$, and the packets' size are used as a set of signal $x(n)$. The samples will be divided into N intervals of length $m(m > 0)$:$I_0, I_1, ..., I_{N-1}$. Within each interval, we can compute its DCT to obtain the DCT coefficients:

$$X(k) = l(k)\sum_{n=0}^{m-1} x(n)\cos\left\{\frac{\pi(2n+1)k}{2m}\right\} \tag{1}$$

With $k = 0, 1, ..., m-1$,where $l(k) = \frac{1}{\sqrt{m}}$ if $k = 0$ and $l(k) = \sqrt{\frac{2}{m}}$ if $1 \le k < m$. Fig. 2 shows the distribution of the DCT coefficient for a certain segment of the signal. Since the energy is more randomly distributed in a large range, simply choose low energy area is very difficult. According to the idea of the mean value of DCT in [14], we are now interested in the "balance energy" of those signal in each interval I_i. We define the center of interval $I_i(i = 0, 1, ..., N-1)$ as:

$$C(I_i) = \frac{1}{m}\sum_{k=0}^{m-1} X(k) \tag{2}$$

The distribution of "balance energy" is presented in Fig. 3.

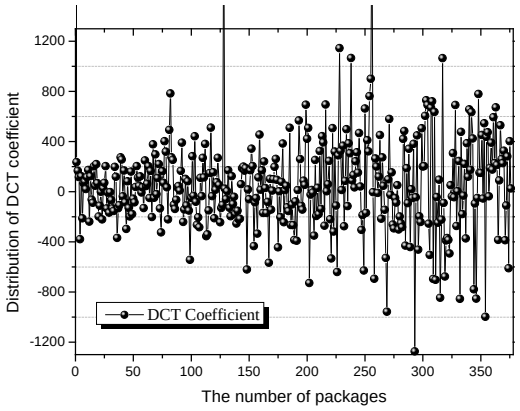

Fig. 2. The distribution of the DCT coefficient

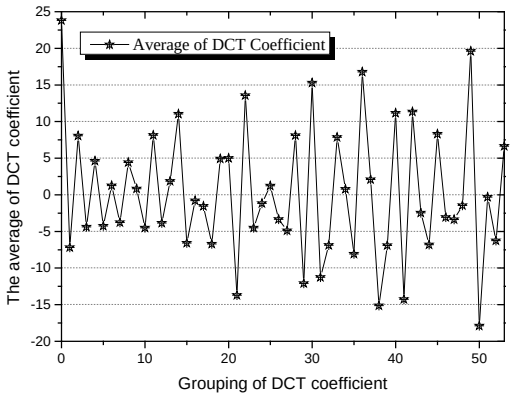

Fig. 3. The distribution of "balance energy"

B. Selecting and embedding

It can be seen from Fig. 3 that the average is more evenly distributed on both sides of the *0*. The closer to zero, the smaller the energy contains in this region. Each point represents a region in the original data stream. We choose those points which locate in *(-f, f)*, then we can get a discrete watermark embedding area in the original stream. With a certain watermark embedding method, the watermarks will be embedded into the flow discretely. At the receiving end, the watermarks can be extracted after denoising.

Experiment

We have made a detailed introduction to the specific application of DCT in the watermarking location in previous sections. In this section, we verify the performance of DCT in improving the robustness and concealment of watermark by simulation experiments.

Simulation Setup. CAIDA data set contains a large part of the data stream from 1 pm. to 2 pm. in Chicago on March 20, 2014, and the general characteristics of the network traffic can be good represented. To simulate a realistic environment, we extract some data stream from the CAIDA dataset as the target flow in the experiment environment. Based on the discrete watermark embedding model, we implemented a simulate environment as shown in Fig. 4 to evaluate the effectiveness of DCTBWL. The watermarking area will be selected by DCTBWL from the data stream sent from the sender, and then we will embed the watermarks in those selected regions by a classical method named double interval centroid-based watermark(DICBW). In order to achieve better experimental effect, more than thousands of packets are used.

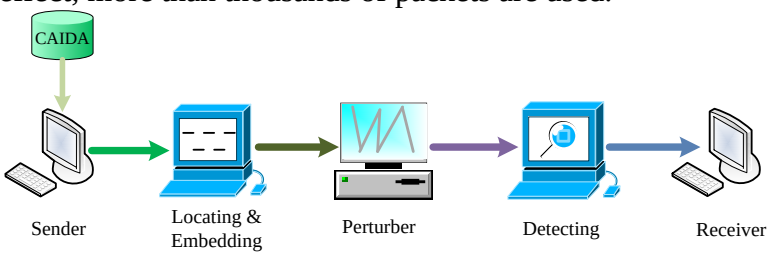

Fig. 4. Abstract model of experiment environment

Influence of Selected Range on Robustness. Since the purpose of the DCT is to choose the watermark embedding positions discretely, the range is very important to the robustness of the watermark. When the range is small, we choose those lower energy area for embedding which are relatively dispersed as is shown in Fig. 3. In this experiment, we use T=600ms, a=90ms in DICBW and 24-bit watermark was embedded at the watermarker. When the selected thresholds ranging from -2.5 to 2.5 and interference rate(Ir) is Ir=0.1, Ir=0.2 and Ir=0.4, the result in Fig. 5 shows that with the increase of watermark redundancy, the detection rates are increased; but the amount of increase is decreased when interference rise. Fig. 6 shows the detection rate under the range of (-5, 5) and (-9, 9) respectively. In the face of the same interference rate, the overall detection rate tended to decrease when the range becomes bigger. The result of the detection when Ir=0.1 is shown in Fig. 7. When all the watermarks can be accommodated, the more dispersed the watermark is, the stronger

its robustness is.

Robustness of DCTBWL Against Interference. The following experiment evaluated the effectiveness of our proposed hybrid watermarking framework in improving the robustness of watermark. In this section, we make the range (-f, f) = (-2.5, 2.5), T = 600ms and a = 90ms. As is shown in Fig. 8, with the same interference rate, the detection rate of the watermark location using DCT is higher than that of the single model. The result demonstrates that both hybrid model and signal model requires at least more than thousand of packets to achieve a high detection rate, while the hybrid model only requires fewer than 1600 packets to be observed to achieve a 100% detection rate when Ir = 0.1. In addition, when the interference rate increases, the gap between the mixed model and the single model becomes larger. Therefor, the DCT positioning can effectively improve the robustness of the watermark in the network transmission.

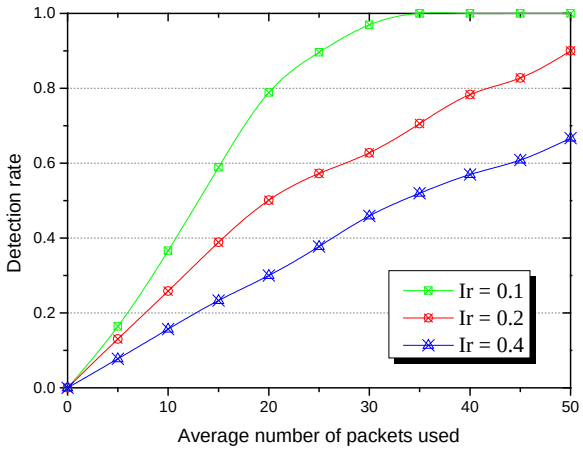

Fig. 5. Detection rate comparison of different interference rate(Ir) ((-f, f) = (-2.5, 2.5), T=600ms, a = 90ms)

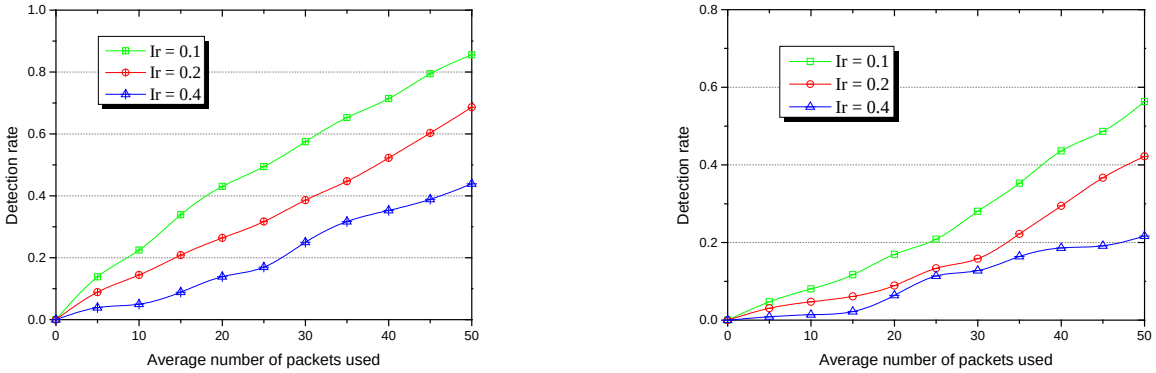

Fig. 6. Detection rate comparison of different selected range (left: (-f, f) = (-5, 5), right: (-f, f) = (-9, 9), T=600ms, a = 90ms)

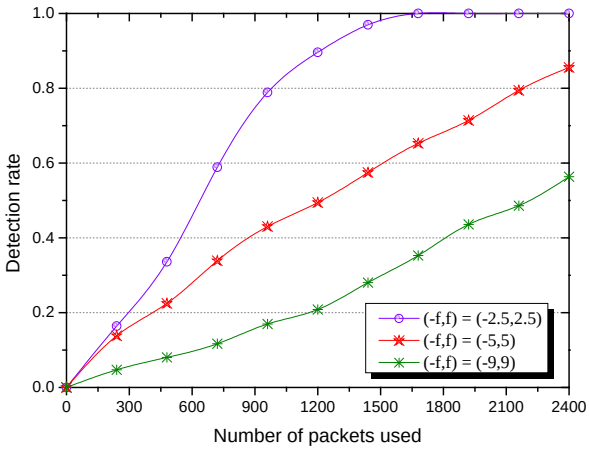

Fig. 7. Detection rate comparison of different selected range (Ir = 0.1, T=600ms, a = 90ms)

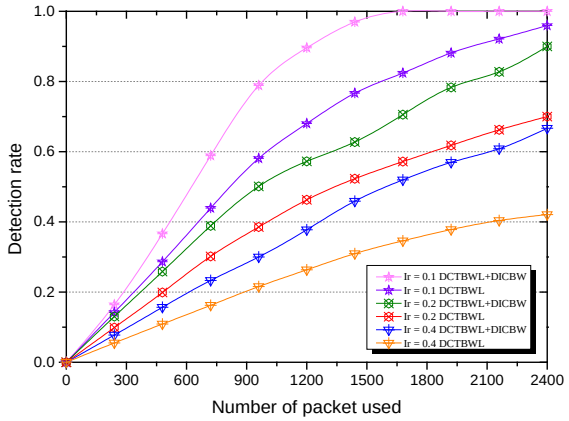

Fig. 8. Detection rate comparison of hybrid model and single model ((-f, f) = (-2.5, 2.5), T=600ms, a = 90ms)

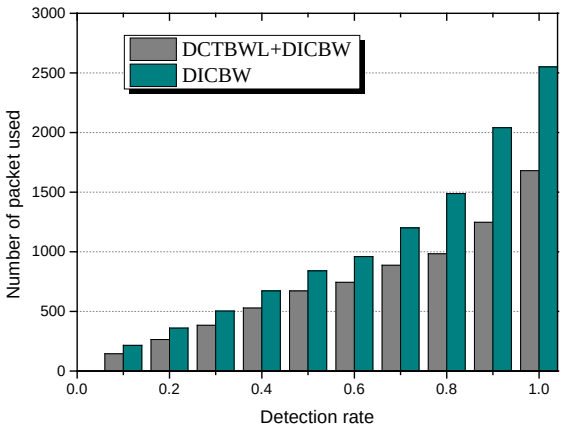

Fig. 9. Comparison of number of packet used between hybrid model and single model (Ir = 0.1, (-f, f) = (-2.5, 2.5), T=600ms, a = 90ms)

The comparison of packet used is displayed in Fig. 9. Under the same disturbing, the method of DICBW needs more packets for embedding. Besides, when accuracy is more than 0.7, with the increase of the detection rate, the number of packets between hybrid model and single model is showing a sharp upward trend.

Conclusion and Future Work

In this paper, we propose a robust positioning method for discrete embedded watermark. By using the Discrete Cosine Transform upon the data stream, we can get the distribution of the energy. After the calculation of the mean, the energy is evenly distributed on both sides of the zero. We can get discrete watermark embedding areas in the original stream by choosing those low energy points, and then with a certain watermark embedding method, the watermarks are embedded into the flow discretely. Experiment results show that the discrete watermark embedding model can reduce the mutual influence between watermarks, improve the detection rate, and increase their concealment. So far, using DCT for location is a new attempt in network watermarking. In the future work, we will consider the issue about improving the detection rate with a high loss rate.

Acknowledgement

This work is supported by National Natural Science Foundation of China (NSFC) under Grant Nos. 61370007, U1405254, 61572206, and Program for New Century Excellent Talents of Fujian Provincial under Grant No. 2014FJ-NCET-ZR06.

Reference

[1] S. Robert, C. Jie, J. Ping, et al. A survey of research in stepping-stone detection[J], International Journal of Electronic Commerce Studies, 2011, 2(2): 103-126.

[2] X.J. Guo, G. Cheng, C.G. Zhu, et al. Progress in research on active network flow watermark[J], Journal on Communications, 2014, 35(7): 178-192.

[3] J. Jiang, J.W. Zhuge, H.X. Duan, et al. Research on botnet mechanisms and defenses[J], Journal of Software, 2012,23(1):82-96.

[4] W. Yu, X.W. Fu, S. Graham, et al. DSSS-based flow marking technique for invisible traceback[A]. Proc of the 2007 IEEE Symposium on Security and Privacy[C]. Oakland, USA , 2007.18-32.

[5] X.Y. Wang, D.S. Reeves. Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays[A]. Proc of the 10th ACM Conference on Computer And

Communications Security[C]. Washington DC, USA, 2003. 20-29.

[6] X.Y. Wang, S.P. Chen, S.S. Jajodia. Network flow watermarking attack on low-latency anonymous communication systems[A]. Proc of the 2007 IEEE Symposium on Security and Privacy[C]. Oakland, USA, 2007.116-130.

[7] X.G. Wang, J.Z. Luo, M. Yang. A double interval centroid-based watermark for network flow traceback[A]. Proc of the 14th International Conference on Computer Supported Cooperative Work in Design[C]. Shanghai, China, 2010. 146-151.

[8] J.Z. Luo, X.G. Wang, M. Yang. An interval centroid based spread spectrum watermarking scheme for multi-flow traceback[J]. Journal of Network and Computer Applications, 2012, 35(1):60-71.

[9] C. Fu, W.Z. Qian, M.Y. Zhao, et al. Delay normalization method of defending against timing-based attacks on anonymous communication systems[J]. Journal of Southeast University (Natural Science Edition), 2009, 39(4):738-741.

[10] X.G. Wang, J.Z. Luo, M. Yang. An efficient sequential watermark detection model for tracing network attack flows[A]. Proc of the 16th IEEE International Conference on Computer Supported Cooperative Work in Design[C]. Wuhan, China, 2012. 236-243.

[11] W.J. Jia, F.P. Tso, Z. Ling, et al. Blind detection of spread spectrum flow watermarks[A]. Proc of the IEEE INFOCOM 2009[C]. Rio de Janeiro ,Brazil, 2009. 2195-2203.

[12] X.P. Luo, J.J. Zhang, R. Perdisci, et al. On the secrecy of spread-spectrum flow watermarks[A]. Proc of the European Symposium on Research in Computer Security 2010[C]. Athens, Greece, 2010. 232-248.

[13] Y. Xiang, I. Natgunanathan, S. Guo, et al. Patchwork-Based audio watermarking method robust to de-synchronization attacks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(9): 1413-1423.

[14] L.Y. Pan, A blind watermark algorithm based on DCT mean[J]. Computer Systems & Applications, 2008, 112-115.