Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

An Interval Centroid Based Spread Spectrum Watermark for Tracing Multiple Network Flows

Xiaogang Wang[1,2], Junzhou Luo[1], Ming Yang[1]
[1]School of Computer Science and Engineering
Southeast University
Nanjing, P.R. China
{wxiaog, jluo, yangming2002}@seu.edu.cn
[2]Changzhou College of Information Technology, Changzhou, P.R. China

***Abstract*—Network flow watermarking schemes have been proposed to trace attackers in the presence of stepping stones or anonymized channels. Most existing interval-based watermarking schemes are ineffective at tracing multiple network flows in parallel due to their interference with each other, while recently proposed Direct Sequence Spread Spectrum (DSSS) watermarking technique is unsuitable for tracing low data rate traffic. By combining interval centroid based watermarking (ICBW) modulation approaches with spread spectrum (SS) based watermarking coding techniques, we herein propose an Interval Centroid Based Spread Spectrum Watermarking scheme (ICBSSW) for efficiently tracing multiple network flows in parallel. Based on our proposed theoretical model, a statistical analysis of ICBSSW, with no assumptions or limitations concerning the distribution of packet times, proves its effectiveness despite traffic timing perturbation and robustness against multi-flow attacks. The experiments using a large number of synthetically generated SSH traffic flows demonstrate that ICBSSW can efficiently trace multiple flows simultaneously and achieve high secrecy by utilizing multiple PN codes as random seeds for randomizing the location of the embedded watermark across multiple flows.**

***Keywords*—watermark, multi-flow traceback, interval centroid based watermarking, spread spectrum**

I. INTRODUCTION

As society grows increasingly dependent on the Internet for commerce, banking, and mission critical applications, network-based attacks are serious threats to the Internet community. Identifying the sources of attacks is a promising approach to holding attackers accountable for their malicious actions and therefore deterring attacks, but it is still a challenging task to trace attackers in the presence of stepping stones [1] or various anonymized channels [2] which render the traditional traceback approach using IP header information ineffective. Recently, various traffic analysis techniques [3] which are the practice of inferring sensitive information from communication patterns have been proposed to break anonymity guarantees and detect stepping-stone intrusion. Among these traffic analysis approaches, timing-based active watermarking [4,5,6,7] can identify the sources of attacks efficiently and are robust to timing perturbation injected by attackers. Traceback is achieved by embedding/decoding watermarks in network flows and correlating the flows with similar watermarks.

Most existing interval-based watermarking schemes subdivide the flow to be marked into discrete time intervals and perform transformative operations on an entire interval of packets. Such schemes are more robust to various timing perturbation than previous approaches that focused on individual packets [1], but they are completely vulnerable to multi-flow attacks by collecting a small number of watermarked flows [8]. Although multiple seed values can be used to watermark different flows for surviving multi-flow attacks, the detection performance is unsatisfied as different flows may interfere with each other significantly when tracing multiple flows in parallel. Recently proposed DSSS watermarking scheme [6] intended to be applied to bulk transfer traffic is unsuitable for tracing low data rate traffic such as web browsing, instant communication and remote login traffic, and it is also vulnerable to a mean-square autocorrelation attack (MSAC) [9] because of the modulated traffic's self-similarity caused by the homogeneous PN code.

Since current watermarking schemes fail to achieve both robustness against watermarking attacks and efficiency of multi-flow traceback, we herein propose a robust and invisible hybrid watermarking scheme (ICBSSW) for tracing multiple network flows simultaneously by combining interval centroid based watermarking modulation approaches (ICBW) [5] with spread spectrum (SS) based watermarking coding techniques. The original watermark is first spread by utilizing a PN code, and then embedded in random preselected intervals by modulating packet timing to withstand multi-flow and MSAC attacks. Redundancy technique is used further to achieve more robustness against various timing perturbation.

The performance of ICBSSW is evaluated based on our proposed theoretical model, with no assumptions or limitations concerning the distribution of packet times. Compared with ICBW, ICBSSW can trace multiple network flows efficiently while being robust against newly proposed watermarking attacks and various flow transformations. The experiments using a large number of synthetically generated SSH traffic flows also demonstrate that ICBSSW can efficiently trace multiple network flows, producing 100% detection rates and less than 1% false positive rates requiring fewer than 1300 packets. Our analytical and empirical results show that ICBSSW can achieve high efficiency when tracing multiple flows simultaneously and high secrecy by utilizing multiple PN codes as random seeds for randomizing the location of the embedded watermark across multiple flows.

The rest of the paper is organized as follows. We review the problem of tracing network flows and existing watermarking

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

techniques in Section Ⅱ. ICBSSW watermarking scheme is presented in Section Ⅲ. In Section Ⅳ, we analyze the effectiveness of ICBSSW based on our proposed theoretical model. We provide experiment results in Section Ⅴ, validating the analysis. The paper is concluded in Section Ⅵ along with some future research directions.

II. RELATED WORK

In order to trace attackers for their malicious conducts, traffic analysis techniques comprising of passive and active approaches have been proposed for anonymous communication traceback and stepping-stone intrusion detection.

The passive traffic analysis approaches make use of various flow characteristics for correlating network flows. Zhang et al. [10] modeled interactive flows as ON–OFF processes and detected linked flows by matching up their ON–OFF behavior. Wang et al. [11] focused on inter-packet delays, and considered several different metrics for correlation. He et al. [12] used packet counts for stepping-stone intrusion detection. Donoho et al. [13] were the first to consider intruder evasion techniques. They defined a maximum-tolerable delay (MTD) model of intruder evasion and suggested wavelet methods to detect stepping stones. Blum et al. [14] used a Poisson model of flows to create a technique with provable upper bounds on false positive rates, given the MTD model. However, thousands of packets are required to be observed for passive traffic analysis approaches to achieve reasonable rates of false errors.

To enhance the efficiency and accuracy of passive traffic analysis, timing-based network flow watermarks utilizing active traffic analysis approaches have been proposed to trace attackers. Wang et al. [1] proposed an active watermarking scheme by manipulation of interpacket delays to detect stepping stones. Fu et al. [15] studied a flow marking scheme to degrade the anonymity in a flow-based wireless mix network by utilizing frequency domain analytical techniques. Based on the distribution of traffic timing, Peng et al. [16] analyzed the secrecy of timing-based watermarking traceback techniques. To make the watermark more robust against flow transformation such as packet losses, insertions and repacketization, Wang et al. [5] proposed Interval Centroid Based Watermarking scheme (ICBW). Pyun et al. [4] proposed Interval-Based Watermarking scheme (IBW) for the purpose of tracing the traffic in the presence of repacketization. To achieve more invisibility, Yu et al. [6] proposed a Direct Sequence Spread Spectrum (DSSS) based traceback technique.

Timing-based active watermarking schemes are effective at tracing attackers, however, the schemes themselves are also under the threats of attackers, and therefore more efforts have been focused on attacking watermarking schemes recently. Kiyavash et al. [8] proposed a multi-flow approach detecting the interval-based or DSSS-based watermarks. In order to make watermarking schemes robust against such attacks, Houmansadr et al. [17] proposed a Multi-flow Attack Resistant Interval Centroid Based Watermarking (MAR-ICBW) scheme by virtue of randomizing the location of the embedded watermark across multiple flows and therefore, effectively removing the correlations between the flows. Since multiple watermarked flows are often aggregated and need to be identified at the detector, the interference caused by different watermarked flows has been a big challenge to existing interval-based watermarking schemes. Recently, Jia et al. [9] proposed a MSAC attack detecting the existence of DSSS-based watermarks, which exploits the mean-square autocorrelation of the traffic rate time series of a single modulated flow. Since current watermarking techniques fail to achieve traceback efficiency, secrecy and accuracy simultaneously, a novel watermarking scheme is required to be robust against newly-proposed watermarking attacks and capable of tracing multiple network flows in parallel.

III. ICBSSW WATERMARKING SCHEME

We propose a hybrid watermarking scheme ICBSSW to achieve robustness against watermarking attacks and efficiency of multi-flow traceback by combining interval centroid based flow modulation methods with SS-based coding techniques.

A. *Watermarking Model*

Figure 1 illustrates ICBSSW watermarking model consisting of a watermark embedding module and a detection module. This layered model can be adopted as a general framework for implementing various watermarking schemes.

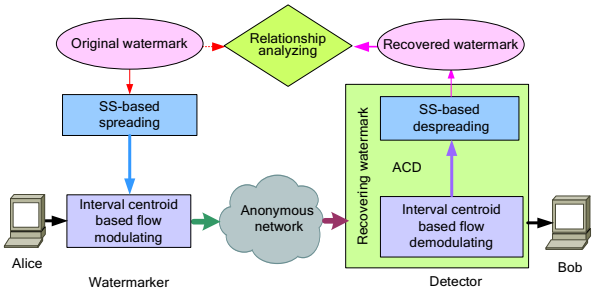

Figure 1. ICBSSW watermarking model

The watermark embedding module at the watermarker is responsible for encoding the original watermark and modulating the target flow. The original watermark is first spread to generate spread watermark by applying SS-based coding technique for achieving high secrecy and accuracy. The spread watermark is then embedded in random preselected intervals by modulating packet timing. The watermarked flow is finally sent through the anonymous network.

The detection module at the detector is responsible for recovering the embedded watermark. The detector first demodulates the received flow to obtain the aggregated centroid difference (ACD) of watermark embedding interval pairs. As the ACD corresponds to the spread watermark, the detector then directly despreads the ACD to recover the watermark using the same PN code shared with the watermarker. Finally, the communication relationship between Alice and Bob can be confirmed by comparing the recovered watermark with the original watermark. If they are equal, the suspected sender and receiver are correlated.

B. *SS-based Watermarking Coding/Decoding Technique*

The spread spectrum (SS) based watermarking technique can achieve high detection rate and low false positive rate.

Spread spectrum is a transmission technique in which a Pseudo Noise (PN) code, independent of the original data signal, is employed to spread the signal over a bandwidth greater than the original data signal bandwidth. At the decoder, the signal is despread using a synchronized copy of the PN code. It is the low cross-correlation of PN codes that makes watermarking invisible and supports multi-flow traceback. Being similar to DSSS technique [6], the basic process of spreading and despreading the watermark is shown in Figure 2.

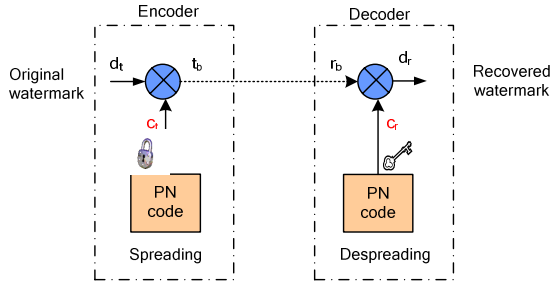

Figure 2. Process of spreading and despreading the watermark

The original watermark to be coded is a series of binary symbols (here we use bits encoded as '+1' or '-1' instead of '1' or '0'). A PN code $\boldsymbol{c_t}$ consisting of '+1' and '-1' is generated at the encoder and shared with the decoder. Each bit $c_{t,i}$ (i=1,…,N_c) in $\boldsymbol{c_t}$ is called a chip, where N_c is the PN code length. To spread one bit d_t, '+1' or '-1' of an original watermark, d_t is directly multiplied with the PN code $\boldsymbol{c_t}$ to produce the spread watermark vector $\boldsymbol{t_b}=d_t\boldsymbol{c_t}$. The received spread watermark vector $\boldsymbol{r_b}$ is despread to recover the original watermark bit by multiplying with a N_c–bit subsequence from the PN code $\boldsymbol{c_r}$ with each bit $c_{r,i}$ (i=1,…,N_c) at the decoder. The recovered watermark bit d_r is calculated as follows:

$$d_r = \frac{\sum \mathbf{r_b} \cdot \mathbf{c_r}}{N_c} = \frac{\sum_{i=1}^{N_c} (r_{b,i} \cdot c_{r,i})}{N_c} = \frac{d_t \sum_{i=1}^{N_c} (c_{t,i} \cdot c_{r,i})}{N_c},$$

where the operator of '·' refers to direct multiplication of vectors and the operator of '∑' adds up all the elements of a vector. If the PN code $\boldsymbol{c_t}$ at the encoder is synchronized with the PN code $\boldsymbol{c_r}$ at the decoder, we have $\boldsymbol{c_r}\cdot\boldsymbol{c_t}=\boldsymbol{1}$, where $\boldsymbol{1}$ refers to a $1\times N_c$ vector with all elements equal to 1. Therefore, we have $d_r=d_t$. If the decoder does not have the right PN code, we have $d_r\neq d_t$. So a decoder without the right PN code cannot reproduce the original watermark bit d_t. The PN code can be considered as a key to recovering the original watermark.

In order to be more robust against flow transformations, the ACD of watermark embedding interval pairs is directly despread to recover the original watermark instead of despreading the spread watermark consisting of a series of discrete binary symbols. As a result, the negative impact of various flow transformations such as timing perturbation added to the ACD can be significantly minimized by virtue of the low cross-correlation of PN codes.

C. Flow Modulation/Demodulation Technique

We assume that the following parameters are pre-distributed: a PN code PN, a random offset o>0, an interval length T>0, an embedding interval selection function S, and a binary watermark w of l bits. The watermark embedding is achieved by slightly adjusting the timing of selected packets within watermark embedding interval pairs. The watermarker embeds the spread watermark w^s into a given flow F. Similarly, given a watermarked flow F^w, the detector detects a watermark and then compares it against the original watermark w for a correlation result.

1) Flow Modulation Technique. All of the time stamps in the redundant r group A and group B intervals ($I_{i,j}^A$ and $I_{i,j}^B$(j=1,…,r)), respectively, are aggregated to modulate spread watermark bit w_i. Let

$$A_i = \frac{\sum_{j=1}^{r}\sum_{k=1}^{N_{i,j}^A}\Delta t_{i,j,k}^A}{N_i^A},\quad B_i = \frac{\sum_{j=1}^{r}\sum_{k=1}^{N_{i,j}^B}\Delta t_{i,j,k}^B}{N_i^B}$$

be aggregated centroids of group A and B packets, respectively, assigned for modulating and demodulating spread watermark bit w_i, where $\Delta t_{i,j,k}^A$ and $\Delta t_{i,j,k}^B$ represent the offset of k-th packet in interval $I_{i,j}^A$ and $I_{i,j}^B$, respectively, satisfying independent and identical uniform distribution in range [0,T). N_i^A and N_i^B are the total number of packets assigned for modulating spread watermark bit w_i in group A and B packets, respectively. Since A_i and B_i are actually the sample means of $\Delta t_{i,j,k}^A$ and $\Delta t_{i,j,k}^B$, both A_i and B_i can be approximated to a normal distribution according to the Central Limit Theorem. The difference Y_i between A_i and B_i is used to modulate and demodulate spread watermark bit w_i, which is symmetric around zero satisfying the normal distribution. Figure 3 illustrates the modulation process of spread watermark bit w_i.

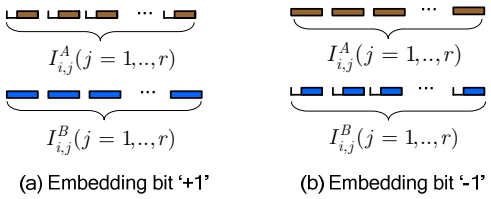

Figure 3. Modulation process of spread watermark bit w_i

To modulate bit '+1', we deliberately increase A_i so that the difference Y_i will be more likely to be positive than negative. Similarly, to modulate bit '-1', we deliberately increase B_i so that Y_i will be more likely to be negative than positive. To increase A_i or B_i, we can simply delay each packet within each interval. Let 0<a<T be the maximum delay to be applied, we delay packet $P_{i,j,k}$ according to the following strategy.

$$\Delta t_{i,j,k}^{'} = a + \frac{(T-a)\Delta t_{i,j,k}}{T} \quad (1)$$

Each watermark embedding interval pair consisting of two intervals is required to modulate 1-bit spread watermark in ICBSSW. The redundancy is implemented by repeating the whole watermark string one after another, thereby dynamically increasing the redundancy for the purpose of adapting to the availability of the packets of the target flow.

2) Flow Demodulation Technique. To extract the spread watermark from a watermarked flow, we can calculate each Y_i (i=1, …, $N_c \times l$) given the correct decoding offset o, interval length T, and the exact interval grouping and assignment information. If Y_i is greater than 0, the demodulation result of spread watermark bit w_i is '+1'; otherwise, the result is '-1'. We assume that the expected number of packets in group A is the same as that in group B, denoted as N_i, for modulating watermark bit w_i given a long lasting target flow. Let Y_i^1 denote the difference of A_i and B_i embedding spread watermark bit '+1', it follows that,

$$E(Y_i^1) = \frac{a}{2},\ V(Y_i^1) = \frac{T^2 + (T-a)^2}{12N_i}. \tag{2}$$

We define the probability that a watermark bit w_i is detected correctly as the detection rate. The detection rate of ICBW without SS-based coding process can be approximated as follows:

$$\begin{aligned}\Pr[Y_i^1 > 0] &= \Pr\left[\frac{Y_i^1 - E(Y_i^1)}{\sqrt{V(Y_i^1)}} > \frac{-E(Y_i^1)}{\sqrt{V(Y_i^1)}}\right] \\ &\approx 1 - \Phi\left(\frac{-E(Y_i^1)}{\sqrt{V(Y_i^1)}}\right) = 1 - \Phi\left(\frac{-a\sqrt{3N_i}}{\sqrt{T^2 + (T-a)^2}}\right)\end{aligned} \tag{3}$$

where $\Phi(\xi) = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$.

Given any interval length T, $0<a<T$, the detection rate can always be increased by increasing N_i, which can be achieved by increasing the redundancy r.

IV. THEORETICAL ANALYSIS

On the basis of watermarking model, we evaluate the performance of ICBSSW by probabilistic methods.

A. Accuracy Under Ideal Conditions

Since the ACD Y_i^1 ($Y_i^1 = A_i - B_i$) denotes the difference of A_i and B_i embedding spread watermark bit w_i^s, multiple values of Y_i^1 (i=1,…,N_c) are directly despread using the PN code $c_{r,i}$ (i=1,…,N_c) to recover the watermark bit w. The random normal variable Y_i^1 can be decomposed as follows $Y_i^1 = \mu_i + v_i$, where μ_i refers to the mean of Y_i^1 and v_i refers to a random variable satisfying a normal distribution with zero mean. Let Y_w^1 denote the result of despreading Y_i^1 (i=1,…,N_c). We have

$$E(Y_w^1) = \frac{a}{2},\ V(Y_w^1) = \frac{T^2 + (T-a)^2}{12N_i N_c}. \tag{4}$$

The detection rate of ICBSSW is expressed as $Pr[Y_w^1 > 0]$ for w='+1' ($Pr[Y_w^1 < 0]$ for w='-1'). According to (4), the detection rate of ICBSSW without any flow transformations can be calculated as follows:

$$\Pr[Y_w^1 > 0] \approx 1 - \Phi\left(\frac{-a\sqrt{3N_i N_c}}{\sqrt{T^2 + (T-a)^2}}\right). \tag{5}$$

The derivation implies that, given any interval length T, $0<a<T$, the watermark detection rate of ICBSSW can always be increased by increasing N_i and the PN code length N_c. We can also increase the detection rate by increasing timing adjustment value a. Note that, given the redundancy r, the number of packets N_i used to modulate spread watermark bit w_i increases as T increases, thereby increasing the detection rate. Compared with ICBW scheme, given the same redundancy r, the watermark detection rate of ICBSSW can always be increased by applying SS-based coding technique regardless of the exact distribution of original traffic.

B. Robustness Against Flow Transformations

In order to evaluate the robustness of ICBSSW, the negative impact of flow transformations on the watermark detection rate should be analyzed. We assume the delay model of Donoho et al. [13] where an adversary is confined to conservative transformation due to the inherent constraint of the interactive traffic. Let R_A, R_B be the ratios between the number of added chaff packets and the number of original packets in group A and B, respectively. By the law of large numbers, we have $R_A \approx R_B \approx R$ when N_i is large. The watermark detection rate of ICBW in the presence of chaff and flow mixing can be approximated as follows:

$$\Pr[Y_i^1 > 0] \approx 1 - \Phi\left(\frac{-a\sqrt{3N_i}}{\sqrt{(T-a)^2 + (1+2R)T^2}}\right). \tag{6}$$

According to (3), the detection rate of ICBW decreases as the R increases. We can always increase the detection rate by having sufficiently large N_i, provided that the flow is sufficiently long and there are enough packets.

To recover the watermark bit w, the ACD Y_i^1 (i=1,…,N_c) is further required to be despread. Let Y_w^1 be the result of despreading Y_i^1 (i=1,…,N_c). When applied to the Central Limit Theorem, the detection rate of ICBSSW can be approximated as follows:

$$\Pr[Y_w^1 > 0] \approx 1 - \Phi\left(\frac{-a\sqrt{3N_i N_c}}{\sqrt{(T-a)^2 + (1+2R)T^2}}\right). \tag{7}$$

Compared with (5) without flow transformations, the detection rate of ICBSSW is decreased due to the timing perturbation. Whereas, compared with ICBW, the detection rate can be significantly increased due to SS-based coding technique despite chaff and flow transformations.

C. Efficiency of Multi-flow Traceback

As Yu et al. [6] argued that tracing multiple flows in parallel is important to achieve efficient traceback, ICBSSW is designed to be able to trace multiple network flows simultaneously. Since different flows may go through the same mix located at the anonymous network, the detector may receive an aggregated flow consisting of multiple watermarked flows. Figure 4 presents a typical scenario for tracing multiple network flows, where three flows (*flow 1*, *flow 2* and *flow 3*) are entering the same mix. The *flow 1* and *flow 3* are aggregated on the output link 2 of the mix. In order to identify each watermarked flow from the aggregated flow, ICBSSW exploits the low cross-correlation of PN codes to address the interference between different watermarked flows. We can use PN code *PN1* to watermark the *flow 1* on input link 1 and PN code *PN2* to watermark the *flow 3* on input link 2. On output

link 2, a detector using copies of these two PN codes, can efficiently identify each flow by applying ICBSSW scheme.

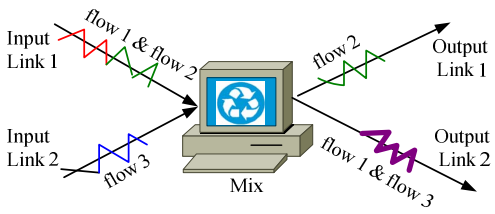

Figure 4. Tracing multiple flows simultaneously

Based on ICBSSW watermarking model, the watermark detection rate of ICBW faced with interference from another watermarked flow can be approximated as follows:

$$\Pr[Y_i^1 > 0] \approx 1 - \Phi\left(\frac{-a(1-R)\sqrt{3N_i}}{\sqrt{(1+R)[(T-a)^2+T^2]}}\right). \quad (8)$$

To recover the watermark bit w, the ACD Y_i^1 (i=1,…,N_c) is further required to be despread. The result of despreading Y_i^1 (i=1,…,N_c), denoted as Y_w^1, can be calculated and the detection rate of ICBSSW is approximated as follows:

$$\Pr[Y_w^1 > 0] \approx 1 - \Phi\left(\frac{-a(1-\frac{R}{N_c})\sqrt{3N_iN_c}}{\sqrt{(1+R)[(T-a)^2+T^2]}}\right). \quad (9)$$

Compared with (5) without any timing perturbation, the detection rate of ICBSSW is decreased due to the interference caused by another watermarked flows. But when compared with ICBW, the detection rate of ICBSSW is obviously increased by SS-based coding technique. Therefore, ICBSSW can effectively identify the watermark of interest from an aggregated watermarked flow, supporting multi-flow traceback.

D. Robustness against Attacks and Watermarking Secrecy

ICBSSW exploits the secrecy of PN codes used as random seeds for RNG to randomly select watermark embedding interval pairs. Since the watermark is embedded by allocating different positions for multiple flows, thereby surviving multi-flow threat attacks based on the alignment approach and the mean square autocorrelation attack based on periodic peaks.

A robust and efficient network flow watermarking scheme needs to be invisible to prevent the watermark from being detected and possibly removed by an active attacker. ICBSSW is invisible to newly proposed attacks by utilizing different PN codes as random seeds for multiple flows. Besides this, the embedded watermark spread by the PN code appears a random message regardless of the original watermark and it is difficult for the attacker to recover the original watermark without right PN code, thereby achieving high secrecy.

V. Experiments

To evaluate the effectiveness of ICBSSW, we implemented a real-time ICBSSW watermarking system shown in Figure 5. As interactive flows are worse scenarios for watermarking schemes than bulk flows, a series of synthetic SSH interactive flows using *tcplib* were generated at Client. A 24-bit watermark w was first embedded at Watermarker. The watermarked flow was then perturbed with random delays at Perturber to simulate the timing perturbation. Finally, the Detector detected a watermark w' from the received flow, which was compared against the original watermark w for a match. In each experiment, the PN code was an m-sequence code of length 7, {1,-1,1,-1,-1,1,1}. We used 50 random flows, each with at least 2000 packets and an average packet rate of 0.86 *packets/second*. These interactive traffic flows exhibited the ON-OFF periods. 75% of the packets had inter-arrival times less than 500ms.

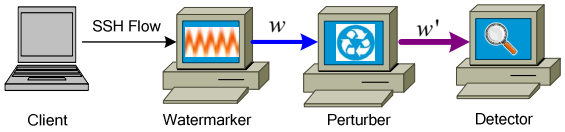

Figure 5. Experiment setup

A. Accuracy Under an Ideal Condition

This experiment measured the detection rate without timing perturbation (D=0ms) by using an interval length T=900ms, and timing adjustment a=500ms, which achieved a redundancy up to r=20, using an average of 800 packets. Detecting thresholds (h) ranging from h=3 to h=7, given the length of watermark ranging from 24 to 32, were used. Figure 6(a) presents the detection rate of ICBW as a function of the redundancy (denoted as r), representing the number of packets used. The result shows that the detection rate increases as the redundancy r (hence the packet usage) increases. The detection rate decreases as the length of watermark increases, which is consistent with the analysis in section Ⅳ. From the result, we can conclude that a higher detecting threshold achieves a higher detection rate for a given redundancy. The detection rate of ICBSSW is presented in Figure 6(b). Compared with ICBW, ICBSSW requires no more than redundancy r=2 for a 100% detection rate, given 24-bit watermark l=24. The result also demonstrates that ICBSSW can achieve the same accuracy as ICBW given the same packet usage.

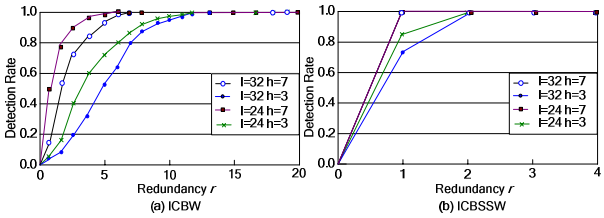

Figure 6. Detection rate comparison of different thresholds (h) (T=900ms, a=500ms, D=0ms)

To evaluate the effectiveness of cross-layer approaches, the detection rate of ICBSSW was evaluated using a cross-layer and our preliminary layered approach [18], respectively. The result shown in Figure 7 demonstrates that ICBSSW using the cross layer approach can achieve higher accuracy than the traditional layered approach. Herein, we adopted the cross-layer approach to implement ICBSSW in the following experiments.

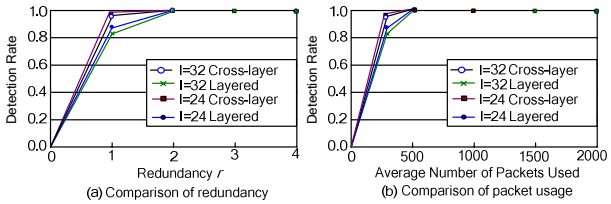

Figure 7. Detection rate comparison between cross-layer and traditional layered approaches for ICBSSW (T=900ms, a=500ms, D=0ms, h=5)

B. *Impact of Timing Perturbation*

The following experiments evaluated the detection rate when faced with active timing perturbation, which was modeled using uniformly random delays with a maximum value of D ranging from 0 to 1200ms. Since the timing perturbation delayed the overall packet timing of the flow, the synchronization process was used at the detector to shift the interval boundaries along with the delayed packets. Offsets were added up with approximately 50% of the maximum delay D. In this experiment, we used an interval length T=900ms, timing adjustment a=500ms, 24-bit watermark l=24 and a detecting threshold h=5. Figure 8 shows the detection rates of ICBW and ICBSSW, respectively, for each different maximum delay D. It shows that, for the same redundancy r, the detection rate decreases as D increases. As stated in theoretical analysis, this occurs because of timing perturbation, which shifts the watermarked packets out of their embedding intervals. As a result, the aggregated centroid difference of watermarked intervals is decreased. Compared with ICBW, ICBSSW only requires the redundancy r less than 2 to achieve a 100% detection rate despite a maximum delay of 400ms. This result highly matches the analytical results. Based on this, we can conclude that ICBSSW is more robust against timing perturbation than ICBW.

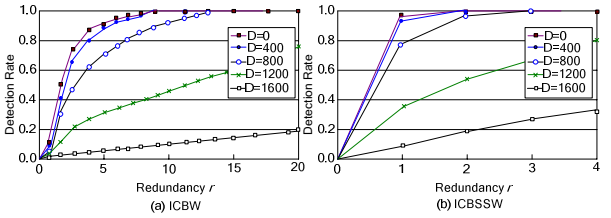

Figure 8. Detection rate comparison of different maximum delays (D) (T=900ms, a=500ms, h=5)

C. *Efficiency and Accuracy of Multi-flow Traceback*

To demonstrate the effectiveness of ICBSSW when tracing multiple flows simultaneously, another watermarked flow generated at Perturber was added to the target flow generated at Client. These two flows modulated using different watermarks were interfered at Perturber, where each watermark had a Hamming distance at least 12 to another watermark. Figure 9 shows the detection rate comparison between ICBW and ICBSSW under the cross-flow interference. The result demonstrates that ICBW requires the redundancy r larger than 40, requiring at least 1500 packets to achieve a 55% detection rate. Compared with ICBW, ICBSSW only requires the redundancy r less than 5 to achieve a 100% detection rate, requiring only less than 1300 packets. This result highly matches the analytical results. We conclude that ICBSSW can effectively and efficiently trace multiple flows simultaneously. The reason is that the watermarked flows are modulated by different PN codes that are minimally cross-correlated. This feature is important for ICBSSW to withstand multi-flow attacks by applying multiple PN codes for watermarking different network flows simultaneously.

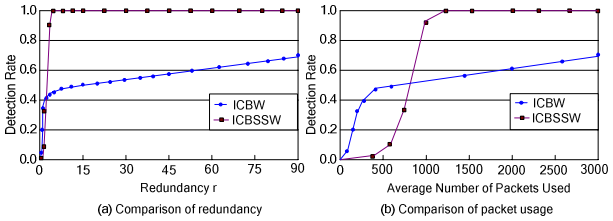

Figure 9. Detection rate comparison faced with cross-flow interference (T=900ms, a=500ms, D=0ms, h=5)

To further investigate the accuracy of different PN code lengths with respect to the cross-flow interference, the detection rate was evaluated using different PN code lengths, ranging from 7 to 31. The result demonstrates that the detection rate increases as the PN code length N_c increases. Note that the choice of N_c exhibits a tradeoff. As N_c increases, the required detecting time also increases due to the increased number of spread watermark bits. The PN code length should be configured concerning performance tradeoff among detection efficiency and accuracy.

D. *Impact of Interval Lengths and Modulation Rates*

We evaluated the detection rate of ICBW and ICBSSW, respectively, using different interval lengths T, ranging from 500ms to 1300ms for a fixed modulation rate 0.5. The result shows that the detection rate increases as T increases. Especially, the performance of ICBSSW is significantly improved over ICBW even the interval length is small (e.g., T=500ms). This feature is essential for secret traceback, which suggests using a smaller T for the sake of secrecy, and it is also important for traceback efficiency, which requires a flow detecting decision to be made over a much shorter time, because of the smaller delays of packets in corresponding intervals.

To investigate the impact of different modulation rates, denoted as the ratios between the timing adjustment values a and interval lengths T, we evaluated the detection rate using different modulation rates, ranging from 0.1 to 0.8, given the interval length T=900ms. The result shown in Figure 10 demonstrates that the detection rate of ICBW and ICBSSW increases as the modulation rate increases. Compared with ICBW, ICBSSW only requires the modulation rate less than 0.4 to achieve a 100% detection rate given the redundancy r larger than 2, which is essential for secret traceback.

To evaluate the false positive rate, we measured the false positive rate both under ideal conditions and timing perturbation with a given 24-bit watermark and threshold h≤5. The percentage of the number of unwatermarked flows that were falsely identified as embedding the given watermark over the 50 flows was measured. In all cases, the respective false positive rates of ICBSSW were measured no more than 1%,

which are less than those of ICBW. The result verifies the effectiveness of ICBSSW.

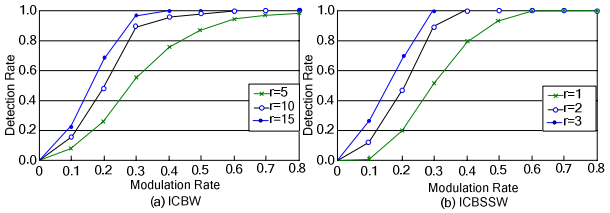

Figure 10. Detection rate comparison of different modulation rates (T=900ms, D=0ms, h=5).

VI. CONCLUSIONS

Watermarking is a promising approach to trace attackers for deterring attacks against electrical commerce. Since existing watermarking schemes fail to achieve both robustness against watermarking attacks and efficiency of multi-flow traceback, the paper presented a theoretical model for the proposed novel watermarking scheme (ICBSSW) which combines interval centroid based watermark embedding methods with spread spectrum based coding techniques. The experiments using synthetically generated SSH traffic flows demonstrated that ICBSSW can efficiently trace multiple network flows, producing 100% detection rates and less than 1% false positive rates requiring fewer than 1300 packets. Our analytical and empirical results show that ICBSSW can achieve high efficiency when tracing multiple flows simultaneously and high secrecy besides withstanding newly proposed watermarking attacks. The performance tradeoffs among robustness, efficiency and secrecy are also identified.

Although ICBSSW provides robustness against flow watermarking attacks and timing perturbation, there are other challenges that have not been addressed in this paper such as splitting or merging of flows. Our future work will investigate the robustness and secrecy of ICBSSW against various potential flow transformations.

ACKNOWLEDGMENT

This work is supported by China Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 200802860031, Jiangsu Provincial Natural Science Foundation of China under Grant No. BK2007708 and BK2008030, Jiangsu Provincial Key Laboratory of Network and Information Security under Grant No. BM2003201, Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under Grant No. 93K-9, and International Science and Technology Cooperation Program of China.

REFERENCES

[1] X. Wang, and D. S. Reeves, "Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays", In *ACM Conference on Computer and Communications Security* (2003), 2003, Washington, DC, USA, pp. 20-29.

[2] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router", In *Proceedings of the 13th USENIX Security Symposium*, 2004, San Diego, CA, USA, pp. 303–320.

[3] A. Back, U. Moller, and A. Stiglic, "Traffic analysis attacks and trade-offs in anonymity providing systems", In *Proceedings of the 4th International Workshop on Information Hiding (2001)*, Lecture Notes in Computer Science, vol. 2848, 2001, Springer Verlag, pp. 245-247.

[4] Y. Pyun, Y. Park, X. Wang, D. S. Reeves, and P. Ning, "Tracing traffic through intermediate hosts that repacketize flows", In *IEEE Conference on Computer Communications* (INFOCOM'07), 2007, Anchorage, Alaska, USA, pp. 634–642.

[5] X. Wang, S. Chen, and S. Jajodia, "Network flow watermarking attack on low-latency anonymous communication systems", In *Proceedings of the IEEE Security and Privacy Symposium* (S&P'07), 2007, Oakland, California, USA, pp. 116-130.

[6] W. Yu, X. Fu, S. Graham, D. Xuan, and W. Zhao, "DSSS-based flow marking technique for invisible traceback", In *Proceedings of the IEEE Security and Privacy Symposium* (S&P'07), 2007, Oakland, California, USA, pp. 18-32.

[7] A. Houmansadr, N. Kiyavash, and N. Borisov, "RAINBOW: A robust and invisible non-blind watermark for network flows", In *Proceedings of the 16th Annual Network & Distributed System Security Symposium*, 2009, San Diego, CA, USA, in press.

[8] N. Kiyavash, A. Houmansadr, and N. Borisov, "Multi-flow attacks against network flow watermarking schemes", In *Proceedings of USENIX Security,* 2008, San Jose, CA, USA, pp. 307-320.

[9] W. Jia, F. P. Tso, Z. Ling, X. Fu, and D. Xuan, "Blind detection of spread spectrum flow watermarks", In *Proceedings of IEEE INFOCOM'09*, 2009, Rio de Janeiro, Brazil, in press.

[10] Y. Zhang, and V. Paxson, "Detecting stepping stones", In *Proceedings of the 9th conference on USENIX Security Symposium*, 2000, Denver, Colorado, USA, pp. 171-184.

[11] X. Wang, D. S. Reeves, and S. F. Wu, "Inter-packet delay based correlation for tracing encrypted connections through stepping stones", In *Proceedings of the 7th European Symposium on Research in Computer Security*, Lecture Notes in Computer Science, vol. 2502, 2002, Springer, pp. 244-263.

[12] T. He, and L. Tong, "Detecting encrypted stepping-stone connections", In *IEEE Transactions on Signal Processing*, 55(5), 2007, pp. 1612–1623.

[13] D. L. Donoho, A. G. Flesia, U. Shankar, V. Paxson, J. Coit, and S. Staniford, "Multiscale stepping-stone detection: detecting pairs of jittered interactive streams by exploiting maximum tolerable delay", In *Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection (RAID'02)*, 2002, Zurich, Switzerland, pp. 17–35.

[14] A. Blum, D. Song, and S. Venkataraman, "Detection of interactive stepping stones: Algorithms and confidence bounds", In 7*th International Symposium on Recent Advances in Intrusion Detection*, Lecture Notes in Computer Science, vol. 3224, 2004, Springer, pp. 258–277.

[15] X. Fu, Y. Zhu, B. Graham, R. Bettati, W. Zhao, "On flow marking attacks in wireless anonymous communication networks", In *Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS'05)*, 2005, Columbus, Ohio, USA, pp. 493-503.

[16] P. Peng, P. Ning, and D. S. Reeves, "On the secrecy of timing-based active watermarking traceback techniques", In *Proceedings of the IEEE Security and Privacy Symposium (S&P'06)*, 2006, Berkeley/Oakland, California, USA, pp. 334-349.

[17] A. Houmansadr, N. Kiyavash, and N. Borisov, "Multi-flow attack resistant watermarks for network flows", In *Proceedings of IEEE International Conference on Acoustics, Speech, and Processing*, 2009, Taipei, Taiwan, in press.

[18] L. Zhang, J. Luo, and M. Yang, "An Improved DSSS-Based Flow Marking Technique for Anonymous Communication", In *Proceedings of International Symposium on Multidisciplinary Autonomous Networks and Systems (MANS 2009)*, 2009, Brisbane, Australia, in press.