INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

ENHANCING INVISIBILITY IN NETWORK FLOWS BY NON-BLIND WATERMARKING TECHNIQUE

[1]Manjumadha.D, [2]Pradeepa.S
[1] M.E. (CSE), Mailam Engineering College, Mailam, India, E-Mail: manjupreethi13@gmail.com
[2]Assistant Professor, Mailam Engineering College, Mailam, India, E-Mail: deepababu.2004@gmail.com

Abstract- Passive traffic analysis can link flows, but requires long periods of observation to reduce errors. Active traffic analysis, also known as flow watermarking, allows for better precision and is more scalable. Previous flow watermarks introduce significant delays to the traffic flow as a side effect of using a blind detection scheme; this enables attacks that detect and remove the watermark, while at the same time slowing down legitimate traffic. We propose the first non-blind approach for flow watermarking, called RAINBOW that improves watermark invisibility inserting delays hundreds of times smaller than previous blind watermarks, hence reduce the watermark interference on network flows. We observe that both RAINBOW and passive traffic analysis perform similarly good in the case of uncorrelated traffic, however the RAINBOW detector drastically outperforms the optimum passive detector in the case of correlated network flows. This justifies the use of non-blind watermarks over passive traffic analysis even though both approaches have similar scalability constraints.

Index Terms- Flow watermarking, Hypothesis testing, Non-blind watermarking, Traffic analysis.

I. INTRODUCTION

In previous work the watermarks introduce significant delays to the traffic flow as a side effect of using a blind detection scheme, this enables attacks that detect and remove the watermark, while at the same time slowing down legitimate traffic. The choice between passive and active techniques for traffic analysis exhibits a tradeoff. Passive approaches require observing relatively long-lived network flows and storing or transmitting large amounts of traffic characteristics.

Watermarking approaches are more efficient, with shorter observation periods necessary. They are also blind rather than storing or communicating traffic patterns, all the necessary information is embedded in the flow itself. This however, comes at a cost. To ensure robustness, the watermarks introduce large delays(hundreds of milliseconds) to the flows, interfering with the activity of benign users and making them subject to attacks.

In this paper we proposed new category for network flow watermarks the non-blind flow watermarks. Non-blind watermarking lies in the middle of passive techniques and watermarking techniques. The non-blind watermarks will record traffic pattern of incoming flows and correlate them with outgoing flows. On the other side, the non-blind watermarking aids traffic analysis by applying some modifications to the communication patterns of the intercepted flows. We develop and prototype the first non-blind watermark, called RAINBOW.

The choice between passive and active techniques for traffic analysis exhibits a tradeoff. Passive approaches require observing relatively long-lived network flows and storing or transmitting large amounts of traffic characteristics. Watermarking approaches are more efficient, with shorter observation periods necessary. They are also *blind*: rather than storing and communicating traffic patterns, all the necessary information is embedded in the flow itself. This however, comes at a cost: to ensure robustness, the watermarks introduce large delays to the flows, interfering with the activity of benign users and making them subject to attacks [16], [17]. Motivated by this, we propose a new category for network flow watermarks, the *non-blind flow watermarks*.

II. PROPOSED SYSTEM

In this approach we use non-blind watermarking which lies in the middle of passive techniques and (blind) watermarking techniques: similar to passive techniques (and unlike blind watermarks), non blind watermarks will record traffic pattern of incoming flows and correlate them with outgoing flows. On the other side, similar to blind watermarks (and unlike passive techniques), non-blind watermarking aids traffic analysis by applying some modifications to the communication patterns of the intercepted flows. We develop and prototype the first non-blind flow watermark, called RAINBOW.

RAINBOW records the *timing* pattern of incoming flows and correlates them with the timing pattern of the outgoing flows. On each incoming flow, RAINBOW also inserts a watermark by delaying some packets, after recording the received timings. As such a watermark is generated independently of flows; this will diminish the effect of natural similarities between two unrelated flows and allow a flow linking decision to be made over a much shorter time period. RAINBOW uses spread-spectrum techniques to make the delays much smaller than previous work. RAINBOW uses delays that are on the order of only a few milliseconds; this means that RAINBOW watermarks not only do not interfere with traffic patterns of normal users, but they are also virtually *invisible* since the delays are of the same magnitude. Traffic analysis is suggested as an effective tool for linking network flows in such scenarios since the intermediate nodes do not significantly modify the traffic patterns of the relayed flows. The common patterns used for traffic analysis are the packet counts, packet timings and packet sizes.

The characteristics of incoming streams and then correlating them with the outgoing ones. The right place to do this is often at the border router of an enterprise, so the overhead of this technique is the space used to store the stream characteristics long enough to check against correlated relayed streams, and the CPU time needed to perform the correlations. In a complex enterprise with many interconnected networks, a connection relayed through a stepping stone may enter and leave the enterprise through different points; in such cases, there is additional communications overhead for transmitting traffic statistics between border routers.

Watermarks improve upon passive traffic analysis in two ways: First, by inserting a pattern that is uncorrelated with any other flows, they can improve the detection efficiency, requiring smaller numbers of packets to be observed and providing lower false-positive rates. Second, they can operate in a *blind* fashion: After an incoming flow is watermarked, there is no need to record or communicate the flow characteristics since the presence of a watermark can be detected independently. The detection is also potentially faster, as there is no need to compare each outgoing flow to all the incoming flows within the same time frame.

III. RAINBOW WATERMARKING

We design a new watermark scheme; we call RAINBOW for Robust and Invisible Non-Blind Watermarking. Our scheme is robust and invisible. However, to achieve invisibility while maintaining detection efficiency, we make the scheme non-blind watermarking. That is, incoming flows timings are recorded and compared with the timings of outgoing flows. This allows us to make watermark test with even low-amplitude watermarks and also non-blind watermarking improves the traffic analysis performance.

In the RAINBOW watermarking scheme, the watermark encoder and decoder share a database which records packet inter-arrival timing. For each individual incoming network flow, the encoder computes and stores the packet inter-arrival timing in DB, then inserts w as extra jitter. RAINBOW records the timing pattern of incoming flows and correlate them with the timing pattern of the outgoing flows. On each incoming flow, RAINBOW also inserts a watermark by delaying some packets, after recording the received timings. As such

watermark is generated independently of the flows; this will diminish the effect of natural similarities between two unrelated flows and allow a flow linking decision to be made over a much shorter time period.
RAINBOW scheme is non-blind, and therefore the detector has access to the IPD database where the unwater marked flows are recorded. Given an observed flow at the detector with IPDs and a previously recorded flow, the detector must decide whether the two flows are linked or not. We also derive the optimum detectors for the RAINBOW watermarks according to the LRT rules. We then derive the optimum passive detectors, showing that the RAINBOW watermark performs significantly better than passive traffic analysis for correlated network flows.

IV. DETECTION TECNIQUES

RAINBOW is the first non-blind watermarking scheme. Non-blind watermarking inherits similar scalability issues from the passive traffic analysis. In this section, we show how non-blind watermarking improves the traffic analysis performance as compared to the traditional passive analysis. We derive optimum likelihood ratio test (LRT) detectors for the RAINBOW watermarking scheme for different traffic models and compare its detection performance to those of optimum passive detectors. We show that RAINBOW outperforms passive traffic analysis for different traffic models; this confirms what we expect intuitively from information theory, as a non-blind watermark detector has access to more information, compared to a passive detector that only has access to the IPDs. We also show that the RAINBOW detector is reliable in different models; we perform our detection analysis for two traffic models:

- Traffic model A: independent flows with i.i.d. inter-packet delays.
- Traffic model B: completely correlated flows.

As it infeasible to evaluate the detection performance for all different traffic models, we discuss the detection performance for these two traffic models and consider any real-world network flow to lie between these two extreme models. We show that an active detector, i.e., RAINBOW, is reliable for different models, while a passive detector fails for certain traffic models.

A. Detection Primitives

We use hypothesis testing to analyze the detection performance of active and passive detectors. For an active detector, we aim to distinguish between the two following hypotheses.

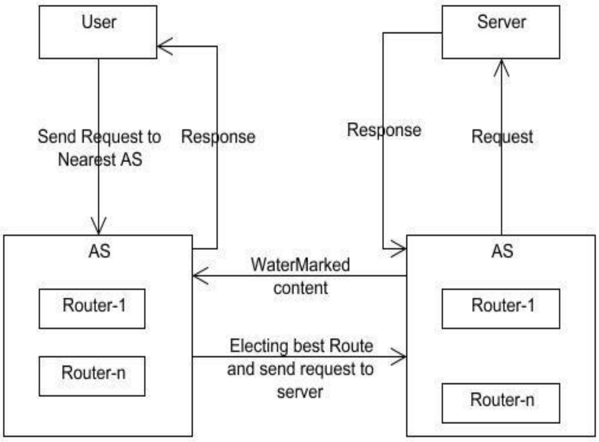

Fig.1: Applying Watermarking Technique

- H_0 (null hypothesis): the received flow with IPDs is a new unwater marked flow, unlinked to the flow with IPDs.
- H_1: is the result of a flow with original IPDs being watermarked and passed through the network.

Also, for passive detector we consider the following hypothesis testing problem.

- H_0 (null hypothesis): The received flow with the IPDs is a new flow, unlinked to the IPDs of another received flow.
- H_1: is the result of passing through the network.

B. *Traffic Model A: Independent Flows, i.i.d.IPDs*

In this model, we assume that the candidate flows are independent. Also, each flow has i.i.d.IPDs, i.e., the flow is modeled with a poisson process. This represents a good model for non interactive network flows.

1.) *Passive Detection:*

In this section, we find the optimum likelihood ratio testing (LRT) passive detector for the traffic model A. Suppose that the flow with IPDs is known to the detector. The detector will need to check if it is correlated with some received flow, where these are independent. Hence, in this case the hypothesis testing problem is,

$$\begin{cases} H_0: & \tau_i^r = \tau_i^* + \delta_i^0 \\ H_1: & \tau_i^r = \tau_i + \delta_i^1 \end{cases}$$

Where these represents the work jitter. Based on our measurements over the planet lab, we model the network jitter with an i.i.d.Laplacian distribution.

2.) *Active Detection:*

In this section, we find the optimum LRT detector for the RAINBOW non-blind watermarked for the traffic model A. We have the following hypothesis testing problem:

$$\begin{cases} H_0: \tau_i^r = \tau_i^* + \delta_i \\ H_1: \tau_i^r = \tau_i + w_i + \delta_i \end{cases}$$

Where these represents the IPDs registered in the IPD database.In order to find the optimum LRT detector, we need to find the distribution of in different hypotheses.

Detection Performance: As before, considering the independence of the IPDs and also the watermarked bits, we use Lemma in the Appendix to find the error probabilities of the ACTV detector for a given T and w,

$$P_{\mathrm{FP}}^{\boldsymbol{\tau},\boldsymbol{w}} \le \prod_{i=1}^{n} e^{-(s\eta_n - \mu_{0,i}^{\tau_i;w_i}(s))}$$

$$P_{\mathrm{FN}}^{\boldsymbol{\tau},\boldsymbol{w}} \le \prod_{i=1}^{n} e^{-((s-1)\eta_n - \mu_{0,i}^{\tau_i;w_i}(s))}$$

C. *Traffic Model B: Correlated Flows, Correlated IPDs*

As the other extreme of traffic models, we investigate the traffic model correlated IPDs, We consider the case where all of the network flows have the same IPDs. This model captures the behavior of a number of widely used types of traffic, including bulk file transfers, browsing, etc. In fact, as we demonstrate in this paper through analysis and simulations, passive traffic analysis is highly invisible.

1.) Passive Detection:

In this model, a passive detection faces the following hypothesis testing problem;

$$\begin{cases} H_0: & \tau_i^r = \tau_i^* + \delta_i \\ H_1: & \tau_i^r = \tau_i + \delta_i \end{cases}$$

Detection Performance: Since the detector is based on random guessing, the false errors are as follows:

$$P_{\mathrm{FP}} = p$$

$$P_{\mathrm{FN}} = 1 - p$$

2) Active Detection:

In this case, we have the following hypothesis testing problem;

$$\begin{cases} H_0: & \tau_i^r = \tau_i^* + \delta_i \\ H_1: & \tau_i^r = \tau_i + w_i + \delta_i \, . \end{cases}$$

Since, $\tau_i^* = \tau_i = C_i$ this can be reduced to the following hypothesis testing;

$$\begin{cases} H_0: & y_i = \delta_i \\ H_1: & y_i = w_i + \delta_i \end{cases}$$

V. OTHER WATERMARK PROPERTIES

A. *Invisibility*

The pioneering design for flow watermarking fails to provide invisibility due to their use of largest packet delays. More recently, Lin *et al.* analyzed watermark invisibility for several flow watermarking schemes, including RAINBOW; they shows that an improper use of watermarking parameters, e.g., large watermark amplitudes, can give away the presence of the watermark. Another analysis was provided by Luo *et al*, where the performance of the scheme was tested against BACKLIT. The authors show that when a watermark is applied only on one side of a TCP connection, it can be detected. To fix this, a watermark should be adapted to be applied on both sides of a connection if the carrying the transport protocol is TCP.

B. *Robustness to Packet Modifications*

A practical watermark detector should withstand packet additions and removals. We showed that RAINBOW resists packet additions/removals up to 20% of the flow length. This is achieved by adding a preprocessing step at the decoder, known as the matching step.

C. *Robustness to Active Attacks*

We note that active robustness is likely to be impossible to achieve simultaneously. This is because to be invisible, a watermarking scheme must introduce small changes to the packet stream. In particular, it cannot introduce jitters exceeding a few milliseconds, as otherwise it would stand apart from the natural network jitter. On the other hand, an active attacker may be willing to introduce large delays. Furthermore, it is easy to imagine an attacker determined to hide the tracks using even more drastic measures. As such, RAINBOW is designed to detect stepping stones when the attackers are unwilling to actively distort the stream as it crosses a stepping stone. As the watermark is invisible, the attacker will not be able to tell that it is being traced and, thus, will be less likely applying watermark countermeasures.

VI. CONCLUSION

In this paper, we introduce the first non-blind active traffic analysis scheme, RAINBOW. Using the tools from the detection and estimation theory, we find the optimum passive and active traffic analysis schemes for different types of the network flows. We show that, for different traffic models, the optimum active detectors outperform the optimum passive detectors. This advantage is more significant for the more correlated network traffic, e.g., the Web Browsing traffic, considering the fact that both passive and non-blind active approaches of traffic analysis are constrained by similar scalability issues, this finding motivated the use of non-blind act iv approaches over the passive approaches.

REFERENCES

[1] X.Wang, S.Chen, and S.Jajodis, “Network flow watermarking attack on low- latency anonymous communication systems”, in *Proc.*IEEES&P, 2007, pp.116-130.

[2] A.Houmansadr, N.Kiyavash, and N.Borisov, “RAINBOW: A robust and invisible non-blind watermark for network flows”, in *Proc.NDDS*, 2009.

[3] A.Houmansadr and N.Borisov,”SWIRL: A scalable watermark to detect correlated network flows”, in *Proc.NDDS*, 2011.

[4] X.Gong, M. Rodrigues, and N. Kiyavash, “Invisible flow watermarks for channels with dependent substitution and deletion errors,” in Proc.IEEE ICASSP,2012,pp. 1773-1776.

[5] A.Houmansadr, “Design, analysis, and implementation of effective network flow watermarking schemes, “ph.D. Dissertation, Dept. ECE, Univ. Illinois at Urbana-Champaign, Urbana, IL, USA, 2012.

.