Dynamic Cross Correlation Matching Algorithm for Digital Video Leakage Detection on Web

Pari Ramalingam B.E., (M.Tech)
Computer Science and Engineering Department, PRIST University, Chennai, India
pariramalingam.mtech@gmail.com

Abstract - **Due to the increased reach of internet, streaming delivery of videos over web is gaining more popularity. One of the major challenges in distribution of video over web is the leakage of video to the external network in the form of redistribution. Peer to Peer (P2P) streaming software is commonly used for the re-distribution of videos over web. Hence the protection of video stream from unauthorized use and duplication is critical for video streaming services. Dynamic Cross Correlation Matching Algorithm compares the traffic pattern from the video streaming server with the traffic pattern from the authorized user who is suspected for re-distribution. It uses a dynamic decision threshold to compare two different video streams.**

Keywords – video leakage; content leakage; leakage detection; traffic pattern; DRM; network dela; packet loss

I. Introduction

With the rapid development of broadband technologies and the advancement of high-speed wired/ wireless networks, the popularity of real-time video streaming applications and services over the Internet has increased by leaps and bounds.

YouTube and Microsoft network video are notable examples of such applications. They serve a huge population of users from all around the world with diverse contents, ranging from daily news feeds to entertainment feeds including music, videos, sports, and so forth, by using streaming transmission technologies.

In addition, real-time video streaming communications such as web conference in intracompany networks or via Internet with virtual private networks (VPNs) are being widely deployed in a large number of corporations as a powerful means of efficiently promoting business activities without additional costs.

A crucial concern in video streaming services is the protection of the bit stream from unauthorized use, duplication and distribution.

One of the most popular approaches to prevent undesirable contents distribution to unauthorized users and/or to protect authors' copyrights is the digital rights management (DRM) technology. Most DRM techniques employ cryptographic or digital watermark techniques.

However, this kind of approaches have no significant effect on redistribution of contents, decrypted or restored at the user-side by authorized yet malicious users. Moreover, redistribution is technically no longer difficult by using peer-to-peer (P2P) streaming software.

On the other hand, packet filtering by firewall-equipped egress nodes is an easy solution to avoid leakage of streaming contents to external networks. In this solution, the packet header information (e.g., destination and source Internet protocol addresses, protocol type, and port number of outgoing traffic) of every streamed packet is inspected. In case the inspected packets do not verify the predefined filtering policy, they are blocked and dropped.

However, it is difficult to entirely prevent streaming content leakage by means of packet filtering alone because the packet header information of malicious users is unspecified beforehand and can be easily spoofed.

The existing proposals monitor information obtained at different nodes in the middle of the streaming path. The retrieved information is used to generate traffic patterns which appear as unique waveform per content, just like a fingerprint.

The generation of traffic pattern does not require any information on the packet header, and therefore preserves the user's privacy. Leakage detection is then performed by comparing the generated traffic patterns.

However, the existence of videos of different length in the network environment causes a considerable degradation in the leakage detection performance.

Thus, developing an innovative leakage detection method robust to the variation of video lengths is, indeed required. In this paper, by comparing different length videos, we determine a relationship between the length of videos to be compared and their similarity.

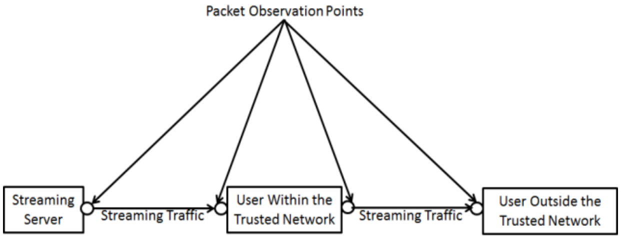
Based on this relationship, we determine decision threshold enabling accurate leakage detection even in an environment with different length videos.

This white paper is intended to answer most of the above queries and help to detect the digital video leakage from trusted networks.

II. Dynamic Cross Correlation Matching Algorithm

This algorithm monitors the traffic pattern information retrieved at different nodes in the network and detects the content-leakage. It focuses on the illegal redistribution of streaming content by an authorized user to external networks. The packet header information (e.g., destination and source

Internet protocol addresses, protocol type, and port number of outgoing traffic) of every streamed packet is inspected.


Packet Observation Points
Streaming Server
Streaming Traffic
User Within the Trusted Network
Streaming Traffic
User Outside the Trusted Network

This algorithm identifies four packet observation points. They are (i) at the streaming server end (ii) at the receiving end of the authorized user (iii) at the sending end of the authorized user (iv) at the receiving end of the malicious user.

The traffic pattern at the above four observation points are captured and compared. The proposed system redefines some of the concepts and models used to characterize traffic patterns. It models the traffic pattern using the PL, PIAT and packet throughput (PT) metrics, with the PT metric being dependent on the PIAT and PL variables.

Typically the video and audio traffic patterns are classified into two types namely CBR and VBR. CBR traffic is defined as "a traffic pattern with a steady bit rate during a time interval"; and VBR as "a traffic pattern with a changing bit rate during a time interval". The proposed system can compare both CBR and VBR.

The algorithm is divided into the following sequence of processes. (i) Pattern Generation (ii) Pattern Matching (iii) Criteria for Leakage Detection (iv) Analyzing the Network Delay, Jitter and Packet Loss for Leakage.

To simulate the working of this algorithm, a content server and a management server was developed. Content server is the one, which stores all video content, which servers the authorized users upon their request for particular content. The server-side traffic pattern is registered and represents the original traffic pattern. Traffic patterns are then generated at the packet observation points.

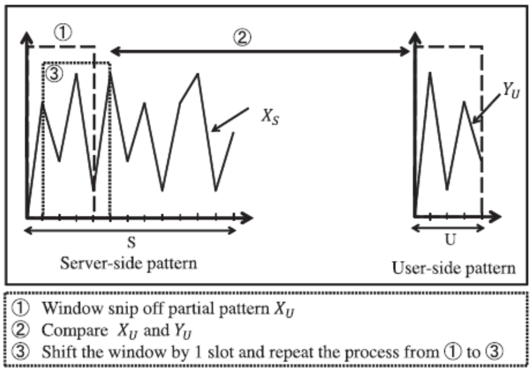
Management server is the one, which monitors the traffic pattern from server side and user side. Time slot based traffic pattern is considered, time slot-based algorithm is a straightforward solution to generate traffic patterns by summing the amount of traffic arrival during a certain period of time.

Authorized user is a user who is authorized to access the content on the content server. Authorized user requests the interested video file to content server. Content server transmits the video to the authorized user by splitting it into number of small chunks/packets. The chunks are transmitted via router to reach the user. The chunks/packets are aggregated at the user side to get the complete video. Authorized user may sometime transfer vide to unauthorized users.

Unauthorized users are one, who gets video file redistributed from the authorized user. They are said to be non-regular users or malicious users. An authorized user in a secured network receives streaming content from a content server and then redistribute it to the unauthorized users.

When the regular user stream video to the non- regular user, it is considered to be content leakage detection. The cross-correlation matching algorithm is performed on the traffic patterns generated through time slot-based algorithm. When there is a variation found, it is detected as content leakage.

The below diagram depicts the approach for comparing the traffic pattern in the server side and the user side (both authorized user and malicious user).


①
②
③
X_S
Y_U
S
Server-side pattern
U
User-side pattern
① Window snip off partial pattern X_U
② Compare X_U and Y_U
③ Shift the window by 1 slot and repeat the process from ① to ③

Y_U = (y1, y2, . . . ; $y_U)^t$. Here, S and U are number of slots, and the length of the user-side observation is shorter than that of the server-side, i.e., S > U.

The fundamental method to quantify the similarity of traffic patterns called cross-correlation matching algorithm, consist of computing the cross-correlation coefficient, which is used as a metric of similarity between the various traffic patterns. Before calculating the similarity between the partial pattern X_U and the server-side pattern Y_U, they are normalized as

$$X'_U = \begin{pmatrix} (x_1 - \overline{x})/s_x \\ (x_2 - \overline{x})/s_x \\ \vdots \\ (x_U - \overline{x})/s_x \end{pmatrix}, \qquad Y'_U = \begin{pmatrix} (y_1 - \overline{y})/s_y \\ (y_2 - \overline{y})/s_y \\ \vdots \\ (y_U - \overline{y})/s_y \end{pmatrix}$$

Here, x and y are the means of each vector. sx and sy are the standard deviations. After normalization, the means and variance of X^1_U and Y^1_U are zero and one, respectively. The cross-correlation coefficient between X_U and Y_U is given by the following equation:

$$R_{X_U Y_U} = \frac{{X'_U}^t Y'_U}{\sqrt{\|X'_U\|^2 \|Y'_U\|^2}}, \qquad -1 \leq R_{X_U Y_U} \leq 1.$$

The conventional approaches, namely, time slot-based traitor tracing (T-TRAT), packet size-based traitor tracing (P-TRAT), and DP-based traitor tracing (DP-TRAT), based on the aforementioned algorithms are summarized in Table 1. The time slot-based pattern generation algorithm used in T-TRAT is influenced by packet delay and jitter, which deteriorate the user-side traffic pattern.

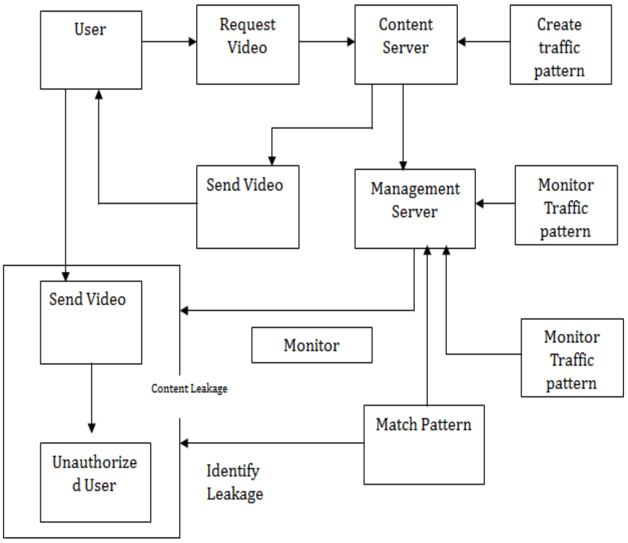
On the other hand, P-TRAT and DPTRAT utilize a traffic pattern generation method based on packet size instead of time slot. As a result, P-TRAT and DPTRAT show robustness against packet delay and jitter. The cross-correlation coefficient is widely use in pattern recognition. However, it is considerably influenced by packet loss that may occur between the streaming server and the user.

Meanwhile, DP matching dynamically alleviates this issue, and shows high robustness to variation in network environment such as the occurrence of packet loss.

The determination of the predefined decision threshold used in P-TRAT and DP-TRAT is done by computing the median between the degree of similarity resulting from the comparison with the same video and the maximum value of the degree of similarity resulting from the comparison with different videos.

A. *Architecture Diagram*

The Dynamic Cross Correlation Matching Algorithm for Digital Video Leakage Detection on Web is designed based on socket programming. Typically the content server runs on a specific computer that has a socket which is bound to a port number. The content server keeps listening in this port number and waits for any users to establish the connection and make the request for the content (video).

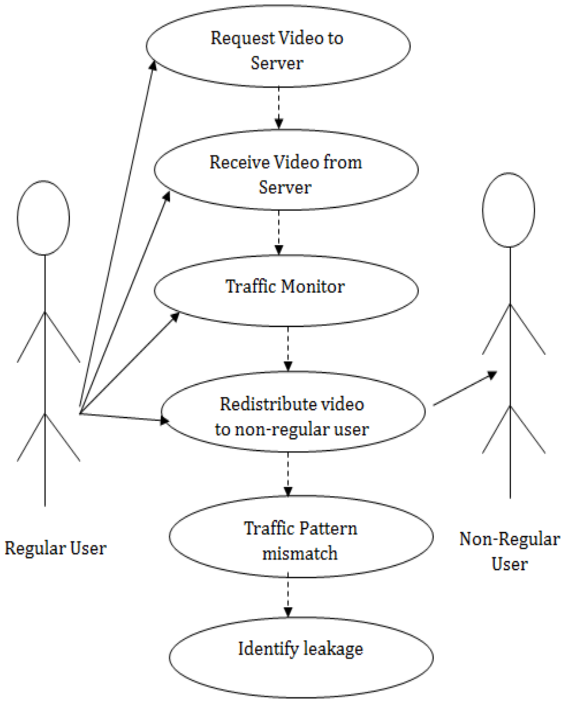


Similarly the Authorized users and the Malicious Users run on different computers which will have sockets which are bound to some port numbers. The authorized user establishes the connection with the content server using the port number on which the content server is listening. He then requests for the content from the content server.

The unauthorized user establishes the connection with the authorized user using the port number on which the authorized user is listening. He then requests for the content from the authorized user.

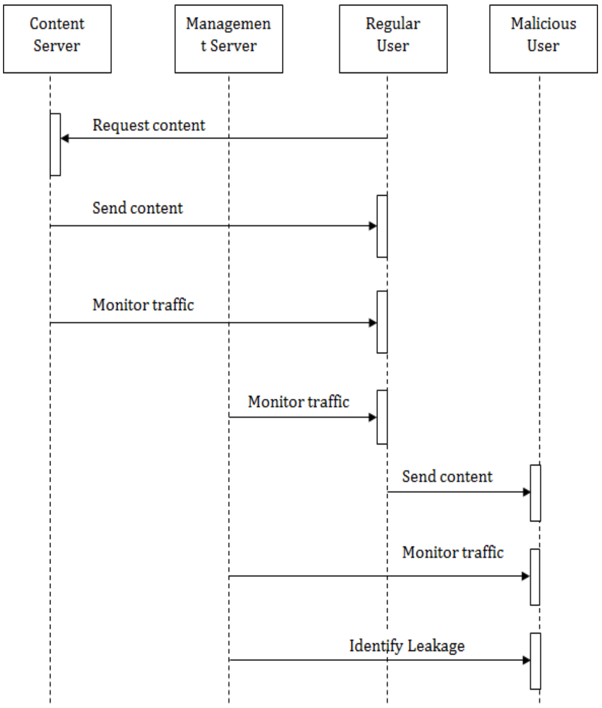
B. *Use Case Diagram*

The below diagram depicts the Use Case flow of the dynamic cross correlation matching algorithm.



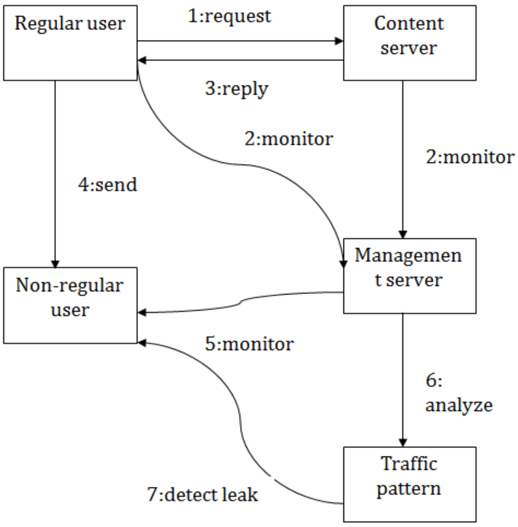
C. *Sequence Diagram*

The below diagram depicts the Use Case flow of the dynamic cross correlation matching algorithm.



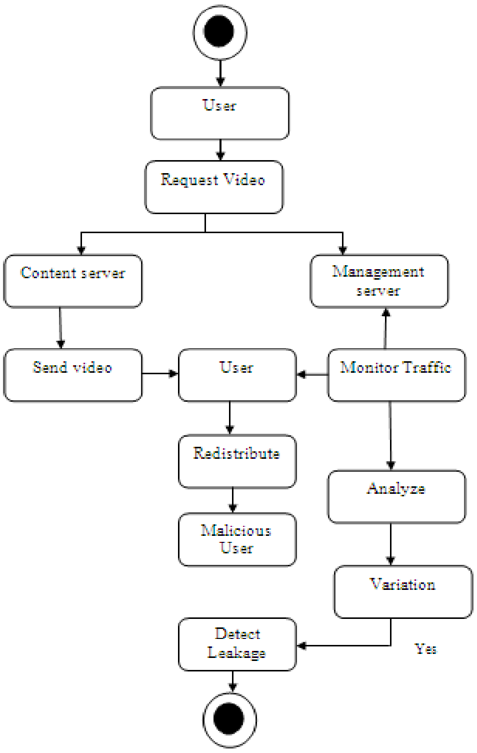
D. *Collaboration Diagram*

The below diagram depicts the Collaboration between the different processes in the dynamic cross correlation matching algorithm.



E. *Activity Diagram*

The below diagram depicts the various activities involved in the dynamic cross correlation matching algorithm.



F. *Abbreviations and Acronyms*

The following abbreviations and acronyms are used in this paper.

VPN – Virtual Private Network

DRM – Digital Rights Management

P2P – Peer-To-Peer

DRM – Digital Rights Management

DP – Dynamic Programming

CBR – Constant Bit Rate

VBR – Variable Bit Rate

PL – Packet Loss

PT – Packet Throughput

PIAT - Packet Inter Arrival Time

TRAT - Traitor Tracing

IDLC - Integrated Data Link Protocol

DDOS - Distributed Denial of Service

III. CONCLUSION

Dynamic Cross Correlation Matching Algorithm for Digital Video Leakage Detection on Web is based on the fact that each of the streaming content has a unique traffic pattern is an innovative solution to prevent illegal redistribution of contents by a regular, yet malicious user.

The uniqueness in the traffic pattern helps to differentiate the regular user and the malicious user. It ensures the high robustness and performance in detecting the leakage even if the network environment changes in terms of the delay, jitter or packet loss.

This algorithm is efficient in detecting the content leakage and is cost effective also. I define the cost in terms of number of matching operations performed in the threshold determination process. The use of the approximation curve enable accurate comparison independently of the length of video.

The performance of this algorithm is also much better than the conventional comparison methods. This algorithm can be used for detecting the content leakage for videos of any lengths. The performance of this algorithm does not degrade even for videos of large lengths.

ACKNOWLEDGMENT

I sincerely express my honour of gratefulness and special thanks to **Mr.A.Mani MCA, M.Tech, (M.B.A),** Assistant Professor, Department of Computer Science and Engineering, PRIST University who has served as an inspiration for me to write this paper.

I also wish to express a special thanks to my mentor **Mr.R.MuthuVenkataKrishnan M.E,** Assistant Professor, PRIST University, for his encouragement and assistance.

REFERENCES

[1] Y. Chu, S.G. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet Using an Overlay Multicast Architecture," Proc. ACM SIGCOMM, pp. 55-67, Aug. 2001.

[2] Z. Yang, H. Ma, and J. Zhang, "A Dynamic Scalable Service Model for SIP-Based Video Conference," Proc. Ninth Int'l Conf. Computer Supported Cooperative Work in DE, pp. 594-599, May 2005.

[3] Y. Chu, S.G. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet Using an Overlay Multicast Architecture," Proc. ACM SIGCOMM, pp. 55-67, Aug. 2001.

[4] O. Adeyinka, "Analysis of IPSec VPNs Performance in a Multimedia Environment," Proc. Fourth Int'l Conf. Intelligent Environments, pp. 25-30, 2008.

[5] E.I. Lin, A.M. Eskicioglu, R.L. Lagendijk, and E.J. Delp, "Advances in Digital Video Content Protection," Proc. IEEE, vol. 93, no. 1, pp. 171-183, Jan. 2005.

[6] S. Craver, N. Memon, B.L. Yeo, and M.M. Yeung, "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications," IEEE J. Selected Areas Comm., vol. 16, no. 4, pp. 573-586, May 1998.

[7] M. Barni and F. Bartolini, "Data Hiding for Fighting Piracy," IEEE Signal Processing Magazine, vol. 21, no. 2, pp. 28-39, Mar. 2004.

[8] K. Su, D. Kundur, and D. Hatzinakos, "Statistical Invisibility for Collusion-Resistant Digital Video Watermarking," IEEE Trans. Multimedia, vol. 7, no. 1, pp. 43-51, Feb. 2005.

[9] E. Diehl and T. Furon, "Watermark: Closing the Analog Hole," Proc. IEEE Int'l Conf. Consumer Electronics, pp. 52-53, 2003.

[10] Y. Liu, Y. Guo, and C. Liang, "A Survey on Peer-to-Peer Video Streaming Systems," Peer-to-Peer Networking and Applications, vol. 1, no. 1, pp. 18-28, Mar. 2008.

Pari Ramalingam received the engineering degree in Computer Science and Engineering from Madras University, India in 1992. He has more than 20 years of experience in Software Industry. He has successfully managed many complex software projects. He is currently pursuing his master's degree in computer science and engineering. His research interests include the areas of network security, mobile computing, software quality assurance and software engineering.