

Wavelet-Based Traffic Analysis for Identifying Video Streams over Broadband Networks

Yali Liu*, Canhui Ou†, Zhi Li‡, Cherita Corbett§, Biswanath Mukherjee¶ and Dipak Ghosal¶

*Department of Electrical and Computer Engineering
University of California, Davis, CA 95616
Email: yliu@ucdavis.edu

†Haas School of Business, University of California, Berkeley, CA 94720

‡AT&T Labs, San Ramon, CA 94583

§Sandia National Laboratories, Livermore, CA 94551

¶Department of Computer Science, University of California, Davis, CA 95616

Abstract—Network and service providers are rapidly deploying IPTV networks to deliver a wide variety of video content to subscribers. Some video content may be protected by copyright and/or may be subject to distribution restrictions. Encryption technologies may not always be effective to manage protected video content, particularly when video content is legally decrypted upon receipt by a subscriber. This paper presents a new approach to detect if specific (or protected) downloaded video is being redistributed by a subscriber using the broadband internet connection. The approach employs a traffic-based signature of the protected video clip. The signature which is shown to be unique is stored in a signature store. We adopt a wavelet-based analysis to match video streams captured from the network to the signatures in the store. The performance of the detection algorithm is evaluated using a large video database populated with a variety of movies and TV shows. The experiment results show that our algorithm achieves high detection rates and low false alarm rates using video clips of only a few seconds.

I. INTRODUCTION

There has been explosive growth in the availability of multimedia content on the web and this content is being distributed over commercial networks. For example, Internet Protocol Television (IPTV) [1] is a system to deliver digital television service to subscribing consumers using the Internet Protocol (IP) over a broadband access network. One obstacle to an effective commercial IPTV network is how to convince content providers that the network can be secured to prevent theft of digital media. For example, to support subscription-based Video-on-Demand (VoD) services, some video content may only be accessible to the paying customers. Therefore, in order to address various legal, copyright, and business issues, it is important to develop robust and efficient methods for monitoring and detecting content leakage.

One general solution is to use encryption algorithms. In typical IPTV services, several levels of authentication and key-exchange procedures are required before the video can be played. However, this may not be effective when video content is legally decrypted upon receipt by a subscriber of a video distribution system. Once the content is decrypted, the subscriber may use a high speed Internet connection to distribute the content to other non-subscribing users. Another approach is digital watermarking [2], in which a watermark

is inserted into the original content prior to distribution so that copies may be traced. However, this scheme is limited by the fact that it significantly increases the complexity of a play-out device and/or reduces video quality. Moreover, in some cases the watermark can be trivially removed. A complementary approach is to detect copies using some inherent properties in the content itself, referred to as content-based copy detection (CBCD) [3]. This approach is based on the paradigm “*the media itself is a watermark*” [4]. By creating unique and robust content signatures, some content leakage can be efficiently traced [5]. Although some deep information about video content, such as colors, shots, textures, can be used to increase the robustness of the video signature, they are limited by the cost of decoding, retrieving, and comparing detailed information.

It has been shown that video traffic has a strong correlation with video content. By examining network traffic information at the IP, TCP, and application layers, video quality degradations can be effectively estimated [6]. In this paper, we propose a new video detection approach based on monitoring video packets in the network. The proposed method does not require decoding the target video sequence to obtain statistical information about each frame. Rather, the signatures are obtained by only analyzing the header of the video packets. The signatures are pre-computed and stored in a database to be used to match with video stream captured from the network. To expedite the detection, any portion of the content can be used based on its availability and speed. In addition, the signature-matching is based on wavelet analysis of the captured video stream to deal with various video traffic impairments from the server or the network, or intentionally introduced by the user to evade detection. Simulation analysis based on various video contents show that our algorithm achieves high detection rates and low false alarm rates using video clips of only a few seconds. More importantly, the proposed scheme is robust against different rate adaptations of video streams.

The rest of the paper is organized as follows. Section II presents a motivating example dealing with detecting video content redistribution over broadband networks. Then, we describe the challenges in signature generation and analysis.

In Section III, we introduce the basic idea of traffic-based signature for video detection. Then we propose a unified Markov model for the traffic modification and a wavelet-based algorithm for matching the signature of a live video stream to a video content. Section IV presents the experimental results based on a MPEG-4 video database with movies and TV shows. Section V concludes the paper and points to future research directions.

II. DETECTING REDISTRIBUTION OF DELIVERED CONTENT

Fig. 1 gives an overview of the content redistribution scenario. Using Set-Top Boxes (STBs) or personal computer with broadband network access (connected to a Central Office (CO) via a Residential Gateway (RG)), video streams may be delivered to a subscriber through an on-demand service request or live broadcast. The downloaded video stream may be stored in a Digital Video Recorder (DVR). Typically, the video encoded with MPEG-4 is delivered to the subscriber using RTP protocol which uses UDP over IP protocols. We consider the case when a subscriber redistributes the downloaded video to other non-subscribing users using the broadband Internet connection. Following are different methods that can be used to redistribute the content. First and the simplest case, the subscriber uses the same protocol stack (same as that is used to download the content) to stream/upload the video. However, the video stream may be rate adapted either due to network conditions or intentionally by the subscribers to evade detection. Second, the subscriber may intentionally alter the video by changing the video coding format (e.g., from MPEG-4 to Microsoft Media Streaming (MMS)) using some transcoding tools or modify the video stream (by adding some noise) to evade detection. Third, the subscriber may store the entire video in a file which can be encrypted, and/or compressed, and uploaded using a different protocol such as HTTP [7] to a server. Fourth, the subscriber may redistribute the content using a P2P protocol such as BitTorrent [8]. Note that in the last case, only a small part of the content may come from one user. In this paper, we only focus on the first case. We assume that the downstream and upstream flows in Fig. 1 use the same protocol stack. However, the upstream rate may be adapted following a general scheme that we will describe in the next section.

In order to detect the content leakage described above, a natural approach is to examine upstream flow from a subscriber and investigate its similarity with the signatures of the monitored video pre-computed and stored in a signature store. We assume that a Deep Packet Inspection (DPI) tool is used to capture the video stream originating from the subscriber as shown in Fig. 1. A wavelet based analysis is applied on the captured video stream which is then matched to the signatures in the store. If the system confirms a match, appropriate actions can be generated (e.g., generate alarm, drop packets, flag offending user for investigation, etc.).

Generally there are two key challenges in the design of a content redistribution detection algorithm. First, we would

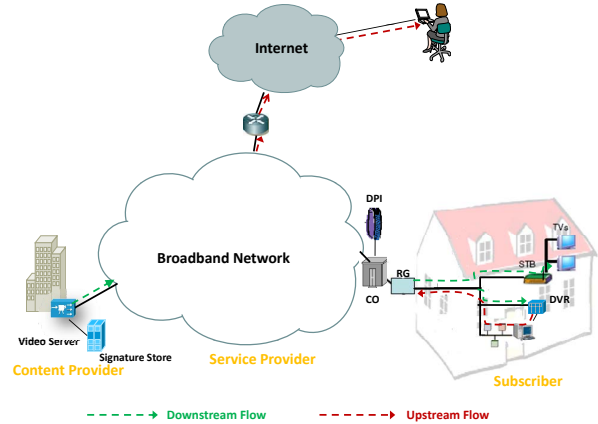


Fig. 1. A content redistribution scenario and detection

like to be able to detect an illegal redistribution in real-time, so that we can respond to the detection before the activity is completed. Therefore, it is important that the scheme be efficient as the comparison may need to be made across a very large database. Second, it is important that the scheme be robust since a redistributed video stream may be impaired by the server or the network in order to handle more connections under heavy loads. Specifically, different users who subscribed to the same movie may see different types of impairments due to large and varying delays, jitter, and loss.

III. WAVELET-BASED ANALYSIS OF VIDEO FLOW

In this section, we first present the method for generating video signatures and discuss the results illustrating their uniqueness. Next, we consider different types of rate adaptations, which can be generalized by a unified Markov model. Finally, we present a wavelet-based matching algorithm to match a video stream with the stored signature.

A. Signature Generation and Uniqueness Analysis

To help optimize the network design for video delivery, an extensive body of literature [6], [9], [10] has investigated and characterized the traffic associated with networked streaming media applications in terms of end-system behavior and network impact. It has been shown that on-line traffic monitoring can create real-time understanding of a particular application and can alert operators to potential performance problems [6].

For IPTV service, we assume MPEG-4 encoded video delivered using RTP, UDP over IP. (Note that while the details of traffic capturing process and pattern would be different for other protocol stacks, e.g., MMS, the basic principles remain the same.) Each RTP transmission starts with a header describing some of the video stream information, e.g. spatial resolution or compression algorithm [11]. We can also obtain the summary information about video frames, including frame type, bytes per frame, and intended times to send and present each frame. The timing information associated with the media frame can be used to determine the position of a captured video

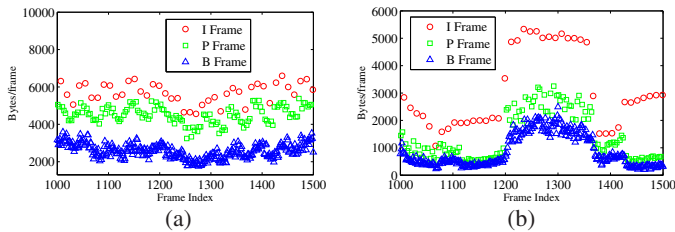


Fig. 2. Video traffic for (a) “StarWarsIV”; (b) “Aladdin”.

clip in the entire video stream. Note that such an examination requires an understanding of streaming protocols, but it does not require examining the content by itself. Thus, we do not need specific knowledge of the video compression algorithms used.

In Fig. 2, we plot frame size vs. frame index from a portion of two video sequences with MPEG-4 video encoder. It clearly shows that, due to their inherent video content, textures, and grain characteristics, different video sequences have distinct traffic patterns for the same small clips. In addition, in the same video clip, there is a high correlation between traffic of successive frames because of temporal correlation and the distinct periodic coding structure as a result of intra (I), predictive (P), and bi-directional (B) coding models. Such traffic behavior is a good indicator of the control, analysis, and performance of content delivery on high-speed networks. Particularly, the unique traffic pattern for each video stream can be used as a signature to trace content redistribution.

It is important to investigate the uniqueness of traffic-based signature since it will determine the feature effectiveness. Among various distance measures for video copy detection, partial Hausdorff Distance (HD) has shown strong discriminatory power in the signature similarity measurement process [12]. Fig. 3 plots the cumulative density function (CDF) of the partial HD (with a partial degree of $f = 0.9$) of any two video clips from 78 different movies and 9 seasons of “Friends” TV shows (total 153 episodes). It clearly shows that the distance between any two video traffic signatures is fairly large. Take TV video clips with the length of 8192 frames as an example. The minimum distance between all the video clips is 2901, and 60% of the distances are more than 5550 frames. Even when the video clip length is as small as 1024 frames, the minimum partial HD between any two video clips is 496. These data show that each video clip has its unique traffic pattern that can be used as the signature for content redistribution detection. We also observe that decreasing a video clip length will decrease its uniqueness. This confirms our intuition that the more information we can use, the more precise information we obtain about the video characteristics.

B. A Unified Markov Model for Traffic Adaptation

In the first case of video redistribution scenario, the upstream video flow depends on the service type (i.e., general users or VIP users), actual video content (i.e., the amount of motion and texture), encoding parameters (i.e., compression

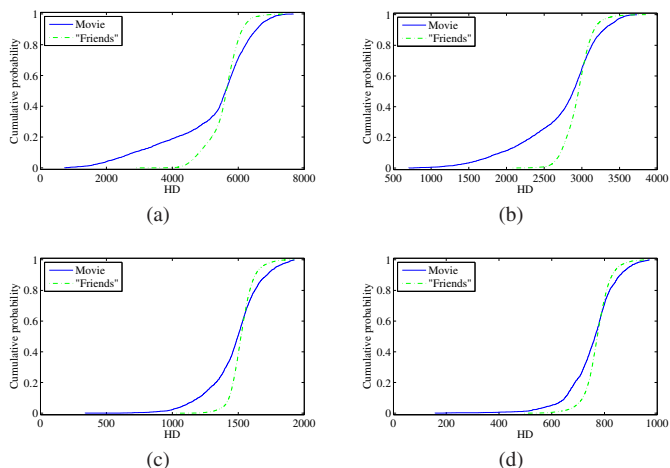


Fig. 3. Cumulative distribution for the partial HD of any two video traffic-based signatures when: (a) $N = 8192$; (b) $N = 4096$; (c) $N = 2048$; (d) $N = 1024$ (N is the video clips length by frames).

algorithm, overall bit rate, spatial and temporal resolution) and any subsequent modifications of the bitstream because of the intentional rate changes by servers in response to channel condition variations or subscribers.

From [6], there are two typical types of rate adaptation. In the first case (Case 1), each video has multiple bit rate streams with different video qualities. When the channel condition changes (e.g., because of congestion), the server decides to reduce the transmission rate by switching from one video stream to a lower-rate video stream at the next I frame. The video stream may roll back to the original or even better video quality at the next I frame if the network or server condition recovers. In the second type of rate adaptation (Case 2), the server may reduce the bit rate by suppressing transmission of B or P frames.

Network congestion or server overload can occur anytime, and in general video coding keeps the I-B-P periodic coding structure. As a result, the rate adaptation process can be modeled as a discrete Markov chain. Fig. 4 illustrates the Markov model for the rate adaptation Case 1. Each state is designated to an integer which is the index of the video quality states. In this case, there are three different video qualities for each video, which are denoted as High (1), Medium (2), and Low (3), respectively. At each I frame, the video bit rate may change to other states or stay at the current state according to the network and server condition. The second rate adaptation scenario corresponds to a two-state discrete Markov model, in which one state denotes dropping the current frame and the other is for keeping the current frame if it is a B or P frame. Therefore, different subscribers who ordered the same video program may get different video stream patterns due to different network conditions and service types (shown in Fig. 5). Especially, in the rate adaptation Case 1, the long-term dependencies of the traffic between adjacent frames may be destroyed but the short-term dependency still remains. On the contrary, for the rate adaptation Case 2, the

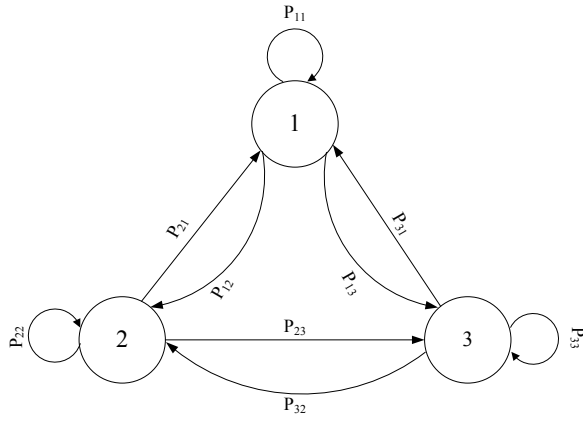


Fig. 4. A unified Markov model for different traffic modifications. (p_{ij} designates the transaction probability from state i to state j)

short-term dependencies may be destroyed but the long-term dependencies are maintained. Considering the practical storage limitation and the number of potential subscribers, however, only limited versions of signatures for each video content may be saved at the signature store. Therefore, a robust matching algorithm is necessary that can exploit both the long-term and short-term dependencies simultaneously.

C. Matching Algorithm

An extensive body of literature suggests that most video traffic can be characterized by self-similar and long-range dependent (LRD) processes [13]. This is a result of the high correlation between successive frames, high motion activities, and the distinct periodic coding structure. However, different rate adaptation schemes may destroy the short-term or long-term dependencies of the video traffic. Consequently, instead of directly analyzing the traffic signal in the time domain,

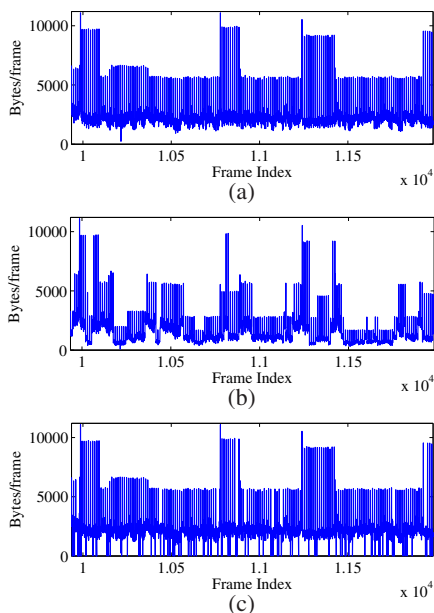


Fig. 5. A video traffic signature from : (a) signature store; (b) a subscriber after rate adaptation Case 1; (c) a subscriber after rate adaptation Case 2.

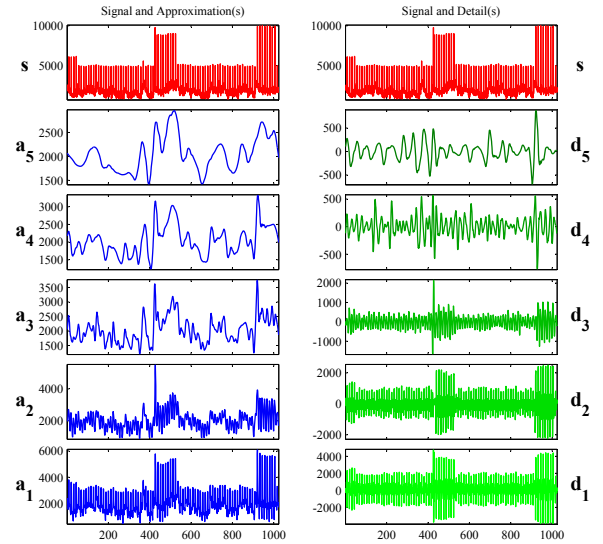


Fig. 6. A video traffic signature s for “Star wars IV” and its wavelet coefficients.

wavelet-based analysis is adopted in our algorithm.

There are two key attributes of wavelet that motivate our choice. First, wavelets allow time-frequency localization, i.e., wavelets allow us to know the characteristics frequency component of the signal and “approximately” where in time they occur. Second, wavelets provide multi-resolution analysis, which allows us to look at the signal at different scales. Fig. 6 illustrates the coefficients of the wavelet transform applied to a portion of video traffic from the Star Wars IV movie. The approximation and detailed coefficients are shown for up to 5 levels.

Given the signature of target video sequence $t(n)$, the goal is to determine if it matches any of signatures $s(n)$ of monitored video streams. Cross-correlation [14] is a standard method of estimating the degree of how two series are correlated. Specifically, the cross-correlation vector of $t(n)$ and $s(n)$ is calculated as

$$c(k) = \frac{\sum_{n=1}^N [t(n) \cdot s(n+k)]}{\|t\| \cdot \|s\|} \quad k = 1, 2, \dots, N \quad (1)$$

where k is the lag of the frame indices of t and s , and N is the signature length. If we detect that there is an obvious peak (outlier) in the cross-correlation vector, the two series will be considered as correlated. One general solution for outlier detection is to use the standard deviation test, in which outliers are treated as samples whose values are τ standard deviations away from the mean. Note that, in our system, the timing information can help to determine the position of the starting points of the captured video clips. Therefore, the outlier should occur at 0 lag position. Fig. 7 shows the cross-correlation vector of wavelet coefficients between the original video traffic and its modified version. Because of the rate adaptation, the traffic patterns of the received bitstream and its original version in the signature store may be different. If we perform the analysis in the time domain, this difference

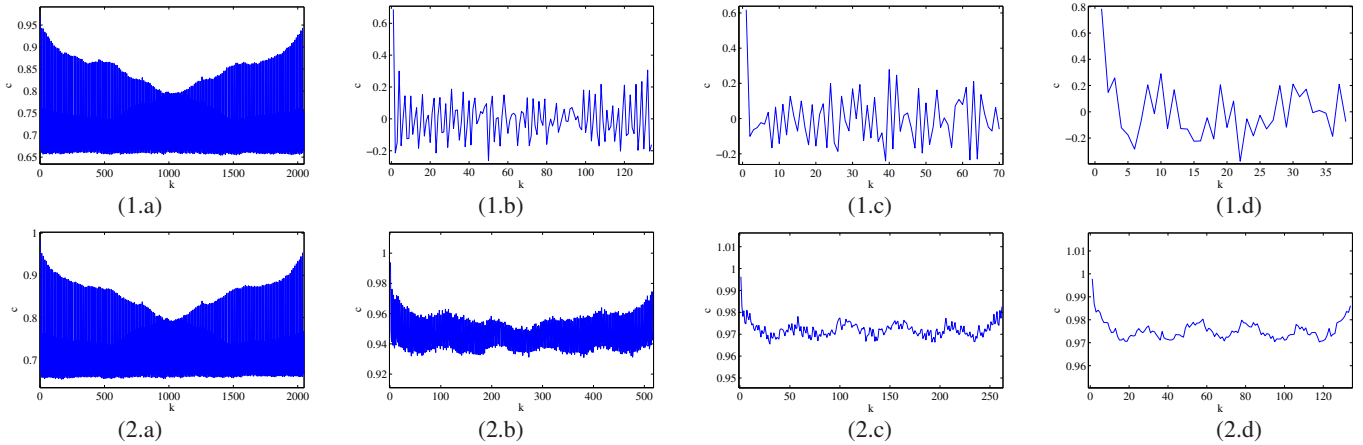


Fig. 7. Matching tests for video traffic signatures and their wavelet coefficients : (1) Rate adaptation Case 1 with (1.a) s , (1.b) d_4 , (1.c) d_5 , (1.d) d_6 and (2) Rate adaptation Case 2 with (2.a) s , (2.b) a_2 , (1.c) a_3 , (1.d) a_4 .

plus the periodic coding structure makes the outlier detection very difficult. However, if the analysis is performed in the wavelet domain, at some wavelet decomposition levels (e.g., level 5 and level 6 in Fig. 7 (1.c) and (1.d)), the outlier is apparent and it can be easily detected. This is because the multi-resolution characteristic of the wavelet transform can help to extract the short-term or the long-term dependencies which are not destroyed by rate adaptation process. In other words, at certain levels, the wavelet coefficients are unique and they can improve the analysis to detect the similarity between the captured video stream and protected video content.

Based of the above reasons, we propose to use wavelet-based multiple matchings to design the matching algorithm. With cross-correlation matching matrix, we can set up a statistical vector \mathcal{I} as:

$$\mathcal{I} = \langle I_n \rangle \quad n = 1, 2, \dots, L \quad (2)$$

to measure the matching result at different wavelet decomposition levels. Here, I_n can only be 1 or 0, which indicates whether two wavelet coefficients match. L is the maximum decomposition level. Its value will be determined by the detector according to the computation cost and detection speed. Thus, the overall matching result \mathcal{V} can be defined as

$$\mathcal{V} = \begin{cases} match & \text{if } \sum_{n=1}^{n=L} I_n \geq \lceil \frac{L}{2} \rceil \\ non - match & \text{otherwise} \end{cases} \quad (3a) \quad (3b)$$

IV. EXPERIMENTAL RESULTS

In our simulation experiments, we use MPEG-4 encoded video streams from public libraries of traces of encoded (compressed) video files [15]. All the traces have been generated from movies of 60 minutes duration or 20 minutes TV shows. For all tests, we use a signature store consisting of 78 movies and 153 TV episodes. The length of captured video clips varies from 8192 (i.e., 300 seconds) to 1024 frames (i.e., 40 seconds) and we set the maximum wavelet decomposition level to $\log_2(N) - 3$, where N is the captured subsequence length. The synchronization process is performed by manually

aligning the starting and ending frames of a captured clip to those of each video clip in the signature store.

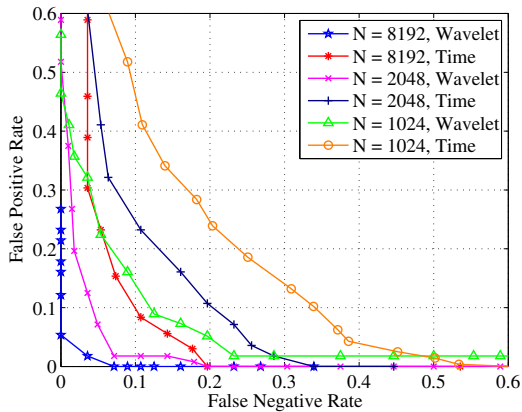
The robustness of the algorithm is measured by receiver operating characteristics (ROC) curve, which is a plot of the false positive rate versus the false negative rate. Each experiment randomly selects 1 clip from 231 videos in the database. Two types of rate adaptation as discussed before are considered. The probability transition matrices are given below (p_1 for Case 1, p_2 for Case 2):

$$p_1 = \begin{bmatrix} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{1}{12} & \frac{1}{6} & \frac{3}{4} \end{bmatrix} \quad (4)$$

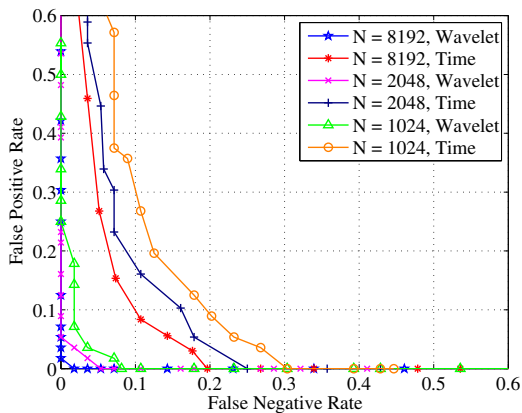
and

$$p_2 = \begin{bmatrix} \frac{19}{40} & \frac{1}{40} \\ \frac{20}{39} & \frac{1}{40} \end{bmatrix}. \quad (5)$$

Different values τ in the matching algorithm generate different false positive and false negative rates shown in Fig. 8; averages computed from 200 independent experiments. The ideal ROC curve should be as close to the axes as possible. Thus we see that, with the help of wavelet analysis, the detection performance dramatically improves. For example, in the rate adaptation case 1, when the length of a captured video clip is about 5 minutes ($N = 8192$), the best operating point has a false positive rate of 0.03 and a false negative rate of 0.01 considering the tradeoff between them. Note that the wavelet analysis using only 40-second traffic information has almost the same performance as that using 300-second traffic information in time domain. This is very useful as the short signatures can reduce the computation cost a lot. In addition, the result for $N = 8192$ has the best performance, followed by those for $N = 2048$ and $N = 1024$. These observations are in accordance with our expectations, as we know that more information about the traffic leads to better detection rates.



(a)



(b)

Fig. 8. Receiver Operating Characteristics (ROC) Curve in: (a) Rate adaptation Case 1; (b) Rate adaptation Case 2.

In addition, we analyze the computation complexity for the proposed algorithm. Here, the complexity is measured by the number of multiplication operations. For a given target sequence with the length N , the number of multiplication operations in the matching test is N^2 for one comparison in the signature store. By using the wavelet analysis, the total number of multiplication operations for all the wavelet coefficients is $(\frac{N}{2})^2 + (\frac{N}{4})^2 + (\frac{N}{8})^2 + \dots \approx \frac{1}{3}N^2$. Particularly, we have shown that $N = 8192$ in time can get the same or even better performance as $N = 1024$ with wavelet. Therefore, compared to direct matching test using time series, the wavelet can help to reduce the computation cost by $\frac{8192^2}{\frac{1}{3} \cdot 1024^2} \approx 200$ times. Since the size of a signature store is very large in real applications, our proposed wavelet-based analysis is a very efficient scheme.

V. CONCLUSION AND FUTURE WORK

We presented a method to detect whether certain video content, such as popular movies, are being redistributed over a broadband network. Particularly, we used traffic information to create video signatures instead of extracting features from deep inside of the content. Considering the possible rate adaptation in the video transmission process which can be modeled with a unified Markov chain, wavelet-based analysis was to determine whether a captured video stream matches

the signatures of a set of video streams in a signature store. We demonstrated through simulation experiments that, using the proposed algorithm, detection can be done efficiently with low false positive and false negative rates.

Some additional questions are yet to be addressed. First, in this work, multiple wavelet coefficients with different wavelet decomposition levels are used to get the long-term and short-term dependency of the video traffic, and then to help deal with different rate adaptation scenarios. One possible future work is to analyze the contribution of different wavelet coefficients. In order to effectively index large databases, the issues of whether we can decrease the number of wavelet coefficients, and which levels should be considered are on the list of our future investigation. Second, a user may evade detection by changing a portion of the video stream or by using other compression techniques, such as winzip, MPEG-4, Real, etc. As a result, further investigation on more powerful distance measures need be conducted to enhance the performance.

ACKNOWLEDGMENT

We would like to express our gratitude to Mr. Raj Savoor of AT&T Labs for his insightful comments which greatly improved the quality of this work.

REFERENCES

- [1] U. Jennehag and T. Zhang, "Increasing bandwidth utilization in next generation IPTV networks," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Oct. 2004, pp. 2075 – 2078.
- [2] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Proc. Second International Workshop on Information Hiding*, 1998, pp. 218–238.
- [3] A. Joly, C. Frelicot, and O. Buisson, "Content-based video copy detection in large databases: a local fingerprints statistical similarity search approach," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2005, pp. I-505–8.
- [4] A. Hampapur and R. M. Bolle, "Comparison of distance measures for video copy detection," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, August 2001, pp. 737–740.
- [5] H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1269–1279, Dec. 1999.
- [6] A. R. Reibman, S. Sen, and J. V. der Merwe, "Network monitoring for video quality over IP," in *Picture Coding Symposium*, 2004.
- [7] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, *Hypertext Transfer Protocol – HTTP/1.1*, June 1999.
- [8] "Bittorrent protocol." [Online]. Available: <http://bitconjurer.org/BitTorrent>
- [9] J. van der Merwe, S. Sen, and C. Kalmanek, "Streaming video traffic: characterization and network impact," in *Proc. Web Caching Workshop*, 2002.
- [10] D. Loguinov and H. Radha, "Measurement study of low-bitrate internet video streaming," in *Proc. of the 1st ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp. 281–293.
- [11] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RTP: A transport protocol for real-time applications*, 1998.
- [12] A. Hampapur and R. Bolle, "Comparison of distance measures for video copy detection," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2001, pp. 737–740.
- [13] R. Grunenfelder, J. Cosmas, S. Manthorpe, and A. Odinma-Okafor, "Characterization of video codecs as autoregressive moving average," *IEEE Journal on Selected Areas in Communication*, vol. 9, no. 3, pp. 284–293, 1991.
- [14] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, "Discrete-time signal processing," *Upper Saddle River, NJ : Prentice Hall*, 1999.
- [15] "Video traces for network performance evaluation." [Online]. Available: <http://www.tkn.tu-berlin.de/research/trace/trace.html>