



Technical University of Crete
Electronic and Computer Engineering Department

Diploma Thesis

Activity-Based Video Traffic Policing: A New Paradigm

Αστυνόμευση της Κίνησης Βίντεο Βάσει Δραστηριότητας: Μια Νέα Προσέγγιση

Maratsolas Evangelos

Advisor: Assistant Professor Polychronis Koutsakis

Chania, November 2012

Acknowledgements

This thesis would not have been possible without the guidance and the help of several individuals, who in one way or another contributed and offered their valuable assistance in the preparation and completion of this study.

First and foremost, my utmost gratitude to my advisor, Professor Polychronis Koutsakis who gave me the possibility to complete this thesis with his supervision, advice and guidance from the very early stage of this research.

I am grateful for his encouragement and precious contribution throughout the elaboration of this study. I would also like to thank him for giving me the opportunity to work on this very interesting field of research. I am indebted to him more than he probably knows.

I would like to thank Assistant Professors Michael Paterakis and Aggelos Bletsas for serving in the examination committee of my diploma thesis. Moreover, I would like to thank my laboratory colleagues for their patience and constructive comments.

Also, I would like to thank all my friends for these great years we spent together and for many wonderful memories.

Most of all, I would like to thank my family for their enormous help, understanding and support throughout all these years as a student.

Abstract

The subject of traffic policing for computer communication networks has been studied extensively in the literature. However, the constant development of new multimedia applications, which are “greedy” in terms of bandwidth and Quality of Service (QoS) requirements, calls for new approaches to the traffic policing problem.

Using as a motivation recent work, which has shown that the classic traffic policing mechanisms provide unnecessarily strict policing for conforming but bursty video users, we propose five simple and efficient new mechanisms, which take into consideration and exploit video activity and the Group-of Pictures (GoP) pattern of the traces. Our proposed mechanisms do not use a token generator based on a fixed rate, but vary the token generation rate according to specific video activity-based algorithms.

The proposed mechanisms show significant improvement in the results for conforming users, and that dynamic traffic policing can provide much higher efficiency than the widely used static mechanisms. Our mechanisms also show that they can provide comparable policing results with the classic mechanisms for non-conforming video users. One of the new mechanisms, the Frame Size Aware Token Bucket (FSA-TB) is shown to clearly excel in comparison to all other policing approaches.

To the best of our knowledge, this is the first time that activity-based video traffic policing is proposed in the relevant literature.

Table of Contents

Acknowledgements.....	2
Abstract.....	3
1. Introduction.....	5
2. H.264 Video Traces.....	7
3. Traffic Policing Mechanisms.....	10
3.1 Existing Mechanisms.....	10
3.1.1 The Token Bucket (TB).....	10
3.1.2 The Jumping Window (JW).....	11
3.1.3 The Moving Window (MW).....	11
3.1.4 The Exponentially Weighted Moving Average (EWMA).....	12
3.1.5 The GoP-Based Token Bucket (GBTB).....	12
3.2 Activity-Based Video Traffic Policing Mechanisms.....	13
3.2.1 The Folded Window (FW).....	13
3.2.2 The Variable Exponentially Weighted Moving Average (V-EWMA).....	14
3.2.3 The Hybrid V-EWMA Token Bucket (VEWMA-TB).....	15
3.2.4 The Frame Size Aware Token Bucket (FSA-TB).....	16
3.2.5 The GoP Modeling Based Jumping Window (GMB-JW).....	17
4. Results and Discussion.....	18
4.1 Results with the use of FW.....	18
4.2 Results with the use of V-EWMA.....	26
4.3 Results with the use of VEWMA-TB.....	32
4.4 Results with the use of FSA-TB.....	35
4.5 Results with the use of GMB-JW.....	37
4.6 Comparison of all mechanisms.....	39
4.6.1 Conforming Users.....	39
4.6.2 Non-Conforming Users.....	42
5. Conclusions.....	44
References.....	45

1. Introduction

Traffic from video services is expected to be a substantial portion of the traffic carried by emerging wired and wireless networks [1-2]. More specifically, according to latest Cisco estimates [24], two-thirds of the global mobile traffic in 2016 will be video traffic. Also, a few months ago, in May 2012, more than 16 billion videos were streamed on Youtube [29]. This explosive growth calls for new sets of traffic control procedures to be implemented in order for the networks to cope with the bursty new applications, which have strict Quality of Service (QoS) requirements. For Variable Bit Rate (VBR) coded video, statistical source models are needed to design networks, which are able to guarantee the strict QoS requirements of the video traffic. Video packet delay requirements are strict, because delays are annoying to a viewer. Whenever the delay experienced by a video packet exceeds the corresponding maximum delay, the packet is dropped, and the video packet dropping requirements are equally strict.

In order to provide the required QoS guarantees, network resources need to be reserved according to both the QoS requirements and the specified traffic parameters of each application. On this subject, one of the fundamental network control issues is the source policing mechanism. The main goal of this control mechanism is to protect the network resources against intentional or unintentional traffic overflow from certain sources. Several policing mechanisms have been proposed in the literature. Traffic that is considered by the mechanisms to be exceeding a user's contract is either dropped immediately or marked, in order to be dropped if needed at any network node, depending on the total traffic load.

Four of the mechanisms which have been most extensively studied (all of them static in nature) are: the Token Bucket and its variations [3- 11, 25]; the Jumping Window [11-13]; the Moving Window (also known as the Sliding Window) [11, 13, 14]; and the Exponentially Weighted Moving Average [11].

In [15], it has been shown that dynamic traffic policing based on accurate H.263 videoconference traffic modeling can clearly outperform the classic static mechanisms, in terms of the percentage of marked packets of conforming users. The reason is that the static mechanisms are unable to cope with the burstiness of video traffic, and hence cause the marking of a significant percentage of the transmitted packets. However, accurate prediction is not possible for all types of video sequences, and even when it is, it often involves a higher degree of complexity (e.g., [20]), which would incur additional computational requirements for the system.

Therefore, in the absence of an accurate video traffic model we need to design traffic policing schemes which can improve the performance of the classic mechanisms. In this work, firstly we study the efficiency of the classic mechanisms for the transmission of H.264 video traffic. Then, we propose and evaluate the performance of five new mechanisms, which take into account and exploit the Group-of-Picture (GoP) pattern of H.264 video traffic, as well as the estimates regarding video sources' behavior based on video activity. By "video activity" we are referring to the sizes of specific video frames which are crucial in order to derive valid expectations about the upcoming video traffic volume. The mechanisms are shown to outperform all the classic mechanisms against which they are compared, in terms of providing much better QoS to conforming video users and comparable policing performance with the classic mechanisms regarding non-conforming video users.

2. H.264 Video Traces

H.264 is the latest video coding standard of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It has recently become the most widely accepted video coding standard since the deployment of MPEG-2 at the dawn of digital television, and it may soon overtake MPEG-2 in common use [17]. It covers all common video applications ranging from mobile services and videoconferencing to IPTV, HDTV, and HD video storage.

Standard H.264 encoders generate three types of video frames: I (intracoded), P (predictive) and B (bidirectionally predictive); i.e., while I frames are intra-coded, the generation of P and B frames involves, in addition to intra-coding, the use of motion estimation and compensation techniques. I frames are, on average, the largest in size, followed by P and then by B frames. The video coding layer of H.264/AVC (Advanced Video Codec) is similar to that of other video coding standards such as MPEG-2 Video. In fact, it uses a fairly traditional approach consisting of a hybrid of block-based temporal and spatial prediction in conjunction with block-based transform coding [17].

In 2007, the Scalable Video Coding (SVC) extension has been added to the H.264/AVC standard. The SVC extension provides temporal scalability, Coarse Grain Scalability (CGS), Medium Grain Scalability (MGS), and SNR scalability in general, spatial scalability, and combined spatio-temporal-SNR scalability [18]. In the rest of this work, we use the term “H.264” to refer to the H.264/AVC video standard.

An important feature of common H.264 encoders is the manner in which frame types are generated. Typical encoders use several Group-of-Pictures (GOP) patterns when compressing video sequences; the GOP pattern specifies the number and temporal order of P and B frames between two successive I frames. A GOP pattern is defined by the distance N between I frames and the distance M between P frames.

In our study on video traffic policing, we have used nine different long sequences of H.264 VBR encoded videos in forty-eight formats, from the publicly available Video Trace Library of [19]. The traces used are in Common Intermediate Format (CIF) (i.e. 352x288 pixels) and in High Definition (HD) 720 and 1080 format (i.e. 1280x720p and 1920x1080i, respectively).

In addition, we used several different Quantization Parameters (QP) for the traces under study. The statistics for the nine traces are presented in Table 1. The first five traces have a GoP size equal to 16, while the next four have a GoP size equal to 12. The length of all the videos is either 10 or 30 minutes. The data for each trace consists of a sequence of the number of bytes per video frame and the type of video frame, i.e., I, P, or B. The interframe period is 33.3 ms. The B3 and B7 encodings denote the number of B frames between successive I or P frames. Hence, e.g., B7 in a trace with GoP=16 means that there is one I and one P frame in the GoP, each followed by 7 B frames.

Video name	Codec	Quantization Parameters	Mean (bits/sec)	Peak (bits/sec)
Tokyo Olympics	B3	16	1625669	14992560
		28	305875	6684000
		38	87645	2710080
		48	24260	488160
Tokyo Olympics	B7	16	1715014	15213360
		28	330236	6801120
		38	92371	2820480
		48	23626	539520
Silence of the Lambs	B3	16	707332	12477840
		28	144317	5567520
		38	42705	1989840
		48	14279	448800
Silence of the Lambs	B7	16	744671	12720480
		28	152234	5706000
		38	43737	2078880
		48	13779	473760

StarWars IV	B3	16	714404	7843440
		28	155670	2520000
		38	46700	1041360
		48	16230	415200
StarWars IV	B7	16	745424	7717680
		28	163573	2561760
		38	48417	1064880
		48	16239	437520
SonyDemo	B3	16	1902281	15898080
		28	384022	6686640
		38	101519	2538960
		48	26609	696960
SonyDemo	B7	16	2004916	15669360
		28	393875	6830400
		38	104736	2627280
		48	26942	764400
Nbc News	B3	16	2964255	14244480
		28	438957	5475600
		38	118458	2364720
		48	32523	846480
Nbc News	B7	16	2975470	14573520
		28	452415	5606160
		38	121284	2429040
		48	31611	893040
Terminator	B2	28	2214602	21722880
		38	701410	7320720
		48	252481	3382800
Kaet's from Mars to China	B2	28	4849710	78457200
Kaet's Horizon	B2	28	1534785	24014640
SonyDemo2	B2	28	2455620	31852080
		38	675377	11956320
		48	228532	4302240

Table 1. Trace Statistics

3. Traffic Policing Mechanisms

In this Section, we first describe, briefly, five static traffic policing mechanisms of the literature which we have used in our study. Then, we proceed to present the five new mechanisms we propose.

3.1 Existing Mechanisms

3.1.1 The Token Bucket (TB)

The token bucket mechanism has been chosen in the recent past as the traffic descriptor for ATM networks, has been widely studied and is currently used in Cisco equipment [28]. The reason for its popularity is its ability to verify easily whether a source conforms to its declared (at call setup) traffic parameters. The token bucket is the predominant method for network traffic shaping.

The basic idea behind the token bucket approach can be described by the following:

- Tokens are put into the bucket at a certain rate. The bucket has a limited capacity.
- Each token represents a permission to the source to send a certain number of bytes into the network.
- After each transmission from the source, tokens, which correspond to the packets transmitted by the source, are removed from the bucket.
- Arriving packets of K bytes are conforming and therefore are immediately processed if there are tokens equivalent to K bytes in the bucket. If the current number of accumulated tokens (i.e., its equivalent in bytes) is less than the corresponding number of packets, the exceeding number of packets is nonconforming.
- Nonconforming packets either wait until the bucket has enough tokens for them to be transmitted or they are discarded or they are marked as nonconforming in order to be discarded in the case of network congestion.
- If no packets wait to be transmitted, tokens can be accumulated up to the size of the token bucket. If the bucket fills with tokens and the source remains inactive or

transmits at a rate lower than the token generation rate, the token buffer overflows and new incoming tokens are discarded, and therefore can not be used by future source packets. In this way the token bucket mechanism imposes an upper limit on the source's burst length, equal to the token bucket size, i.e., a token bucket permits burstiness, but bounds it. This bound can be described by the following formula:

$$A(s,t) \leq \sigma + \rho(t-s), \quad s < t \quad (1)$$

where $A(s,t)$ denotes the amount of traffic leaving the bucket between times s and t , σ is the maximum burst size and ρ is the token generation rate.

3.1.2 The Jumping Window (JW)

The Jumping Window mechanism uses windows of a fixed length T side by side through time. A new window starts immediately after the conclusion of the previous one. During a window, only K bytes (or packets) can be submitted by the source to the network. In the case that a source attempts to transmit more than K bytes, the excessive traffic is dropped (or marked as nonconforming, as in the case of the Token Bucket). The mechanism is implemented with the use of a token counter, similar to the one of the Token Bucket, and in each new window the associated packet counter is restarted with an initial value of zero [11].

3.1.3 The Moving Window (MW)

The Moving Window (Sliding Window) mechanism is similar to the Jumping Window, but more stringent and more complex to implement. This mechanism again ensures that the maximum number of bytes transmitted by a source within any given time interval of duration equal to the fixed window size, T , is upper bounded by K bytes.

The difference with the Jumping Window mechanism is that each video frame size is remembered for the width of exactly one window, starting with the specific video frame and ending T frames later. This mechanism can be interpreted as a window, which is steadily moving along the time axis, with the requirement that the frame sizes of T frames are stored for the duration of one window [11]. This is the

reason that the implementation complexity is considerably higher than for the other two mechanisms (Token Bucket and Jumping Window), as the complexity is directly related to the window size; also, since the content of successive time windows differs by just one frame, it is clear that the mechanism enforces the strictest bandwidth enforcement policy compared to the Token Bucket and the Jumping Window mechanisms.

3.1.4 The Exponentially Weighted Moving Average (EWMA)

The EWMA mechanism uses consecutive-time windows like the JW mechanism. The difference is that the maximum number of accepted packets in the i -th window (N_i) is a function of the allowed mean value of the video trace per interval N and an exponentially weighted sum of the number of accepted packets in the preceding intervals according to the rule

$$N_i = [N - g * S_{i-1}] / (1 - g), \quad 0 \leq g < 1, \quad \text{with } S_{i-1} = (1-g) * X_{i-1} + g * S_{i-2} \quad (2)$$

which can also be expressed as

$$N_i = [N - (1-g) * (g * X_{i-1} + \dots + g^{i-1} * X_1) - g^{i+1} * S_0] / (1-g) \quad (3)$$

where S_0 is the initial value of the EWMA measurement.

The factor g controls the flexibility of the algorithm with respect to the burstiness of the traffic. If $g=0$, N_i is constant and the algorithm is identical to the JW mechanism. A value of g greater than 0 allows more variable source behavior. Although the computation of N_i can be made efficient for special values of g , the implementation complexity of this mechanism is slightly higher than that of the previous mechanisms.

3.1.5 The GoP-Based Token Bucket (GBTB)

This mechanism was proposed in [25]. It is especially tailored for video traffic, of any GoP pattern. By taking into consideration the GoP pattern, GBTB uses 3 different token buckets, one for each type of video frame (I, P, B). Depending on the

type of the video frame, which is expected to arrive at any given instance (this is known from the GoP pattern), the respective token bucket is activated.

The mechanism can also be implemented with one token bucket, which uses different token generation rates depending on the expected video frame arrival. GBTB was an attempt to solve the problem of the poor performance of all four mechanisms presented in Sections 3.1.1-3.1.4, in the case of conforming but bursty video users. By exploiting the GoP pattern of each trace, GBTB was shown to largely outperform the standard TB mechanism, despite the fact that the bucket size used in GBTB was much smaller.

3.2 Activity-Based Video Traffic Policing Mechanisms

The idea proposed in GBTB, i.e., to exploit knowledge about video sources' behavior in order to improve traffic policing for conforming users, is a step towards a new paradigm in video traffic policing. However, GBTB is again static in nature, as it makes no effort to utilize any other knowledge regarding the video source besides the GoP pattern. In this thesis, we take this idea further and propose new traffic policing mechanisms which exploit knowledge and estimates based on video activity.

3.2.1 The Folded Window (FW)

The Folded Window mechanism is the closest, conceptually, to GBTB, among the five mechanisms we proposed and evaluate. The idea is the following: if the Moving Window is implemented with any window size equal to the GoP or larger than the GoP, then at least one I frame will be included in the "policed" frames. This is not a good choice, because we are forcing the policer to handle together frames of very different sizes (typically, I frames are much larger than P and B frames). What is even worse, in such an implementation, is that the I frame included in the policing of the P, B frames of the n^{th} GoP will be the I frame of the $(n+1)^{\text{th}}$ GoP, which may significantly differ in size in comparison with the I frame of the n^{th} GoP (e.g., in the case of a scene change). This can lead to unnecessary and undesirable strictness or

leniency of the mechanism. Hence, we propose to implement the Moving Window by "folding" the window in two, and using $GoP/2$ as the window size. In this way, half of the examined windows contain an I frame and half of them do not. Next, we proceed to generate tokens not by using the mean rate of the source (as in the Moving Window) but by using two different token generation rates, each depending on the number of I,P,B frames in each window. For example, in the case of $GoP=12$, B2 encoding, half of the windows will contain 2 P and 4 B frames and will have a token generation rate equal to two times the mean P frame size plus 4 times the mean B frame size; the other half of the policed windows will contain 1 I, 1 P and 4 B frames and will have a different token generation rate, computed again based on the number of each type of frames in the window.

The reasons that we don't use an even smaller moving window size than $GoP/2$ is that such a choice would make the already strict MW-type mechanism even stricter, and would increase the computational complexity through the existence of multiple different token generation rates. The use of $GoP/2$ window sizes is enough to separate windows including I frames from those that do not.

3.2.2 The Variable Exponentially Weighted Moving Average (V-EWMA)

The V-EWMA mechanism, which is implemented with three different window sizes (GoP , $GoP/2$, $GoP/4$) is a modified version of the EWMA mechanism. The difference between the two mechanisms is that V-EWMA uses variable values of g , depending on the video activity. More specifically, the idea is to adjust the strictness of the mechanism depending on the sizes of the frames of the past few windows; this choice helps us make a prediction on the expected sizes of the frames of the immediate next windows. Hence, by changing the value of g we are able to control the strictness of the mechanism. The algorithm we implemented for V-EWMA is the following:

```

if  $N_i, N_{i+1} < \text{mean frame size in a window}$ 
    if  $N_i > N_{i+1} > N_{i+2}$ 
         $g = g + 0.05$ 
    end
else if  $N_i, N_{i+1} > \text{mean frame size in a window}$ 
    if  $N_i < N_{i+1} < N_{i+2}$ 
         $g = g\_initial$ 
    end
end
end

```

The logic of the algorithm is that, for a given constant window size, if three consecutive windows contain frames with sizes smaller than the mean, and the frame sizes in each subsequent window are smaller, then it is highly probable that a window with large frame sizes will soon arrive. For this reason, we increase the value of g , in order to make the mechanism more lenient. If, on the other hand, frame sizes in subsequent windows grow larger while they are already above the mean, then for a conforming user we expect soon the arrival of windows with small frame sizes; hence, there is no reason for increased leniency, so we let g return to its initial (default) value. This latter policy is also useful in the case of a non-conforming user, in order to prevent him from continuing to transmit at high rates.

3.2.3 The Hybrid V-EWMA Token Bucket (VEWMA-TB)

This mechanism, as indicated by its name, is a combination of the V-EWMA and the Token Bucket mechanisms. More specifically, for P and B frames we use the classic TB mechanism, while for I frames we use VEWMA in order to offer the system the ability to dynamically adjust to the large changes in I frames' sizes, which usually reflect respective changes in the video scene activity.

3.2.4 The Frame Size Aware Token Bucket (FSA-TB)

The FSA-TB mechanism uses the basic idea of GBTB (three different token buckets, one for each type of video frame), but builds upon it by introducing a video activity-based policing component. More specifically, a well-known and extensively studied fact (e.g., [27]) regarding video traffic is that traffic of the same video trace can be divided into high and low activity scenes. Therefore, when an I frame arrives that is much larger/smaller than the mean, it is expected that the P and B frames of the same GoP will exhibit similar behavior, i.e., the I frame is representative of the activity in the whole GoP. Hence, in FSA-TB we calculate the ratio

$$R = \text{allowed_size}(I_frame) / \text{mean_size}(I_frames) \quad (4)$$

at the beginning of each GoP, and then we use R to adapt, respectively, the token generation rate for the P, B frames of the same GoP as follows:

$$\text{token_generation_rate_P} = R * \text{mean_size}(P_frames) \quad (5)$$

$$\text{token_generation_rate_B} = R * \text{mean_size}(B_frames) \quad (6)$$

This approach is intuitively very well-tailored to the needs of conforming users, as it is based upon the idea of dynamically adhering to the users' instantaneous needs and not trying to police them statically based on a generic mean rate. This approach also takes into account the case of non-conforming users, which remains the major role of any policing mechanism: the ratio R is based on the *allowed* size of each I frame, i.e., when a user tries to send I frames which cannot be transmitted in full (insufficient number of tokens), only the portion of the traffic that can be transmitted is entered into the computation of R ; if the whole size of the I frame was used in the numerator of R , then the user could potentially be allowed to send very large P and B frames, without any substantial policing.

The efficiency of FSA-TB for both the cases of conforming and non-conforming users will be shown in our results, in Section 4.

3.2.5 The GoP Modeling Based Jumping Window (GMB-JW)

The GMB-JW mechanism is rather different, conceptually, than the other four mechanisms we propose in this thesis. It shares with them the idea of performing policing based on information at the GoP scale (i.e., not with the classic approach of using the generic traffic parameters, such as the mean and the peak), but it uses a coarse-grain video traffic model, proposed in [26] for estimating the size of a GoP (in bytes). Of course, many finer-grain models have been proposed in the literature, even within the same work in [26]. However, as we explained in the Introduction, the use of accurate modeling, which was shown in [15] to provide excellent traffic policing results, may not be possible for various reasons. Hence, a coarse-grain estimate of the GoP size needs to be evaluated as an alternative. The simple, first-order Markov chain model proposed in [26] consists of M states, where M is equal to $\frac{G_{\max}}{\sigma_G}$.

G_{\max} denotes the size of the largest GoP and σ_G the standard deviation of the GoP sizes. Thus, the size of the quantization interval is σ_G . The entries of the

transition probability matrix $\{P_{ij}\}$ are estimated by $P_{ij} = \frac{n_{ij}}{\sum_{k=1}^M n_{ik}}$, where n_{ij} denotes

the number of transitions from interval i to interval j .

Our proposed mechanism, GMB-JW, uses a window size equal to GoP and operates similarly to the JW, with the difference that the token generation rate is determined dynamically by the Markov chain model, instead of being constant and equal to the video user's mean declared rate.

4. Results and Discussion

Our simulations were conducted in a Matlab environment, with an Intel Core 2 Duo, 2.4 GHz processor. We implemented all of our proposed mechanisms, together with the classic mechanisms from the literature.

We first compare the conceptually similar new and classic mechanisms in couples, and then we compare all mechanisms in terms of their policing efficiency both for conforming users (the problem at the core of this work) and non-conforming users. Because of the large number of video traces used in our study, we have collected an equally large number of results. To avoid repetition, we present only representative results from each comparison.

4.1 Results with the use of FW

We compare the Folded Window with the Moving Window mechanism. Regarding the MW, we started with a very strict policy, controlling the source transmission rate per video frame (allowing the source to transmit no more than the mean rate in every frame), and progressively we increased the window size, up to 100 video frames. The percentage of marked traffic decreases as the window size increases, for all the traces.

As shown in Figures 1-9, the results of the FW are much better than those achieved with MW, for conforming users. The MW mechanism needs to be implemented with very lenient parameters (windows larger than 50, 60 or even 80 frames in some cases) in order to be able to achieve the same results that the FW achieves with the GoP/2 window size (equal to 6 or 8 frames, depending on the trace). The conclusions derived by the Figures are further confirmed by the results presented in Table 2, where we make a head-to-head comparison between FW and MW for a window size equal to GoP/2.

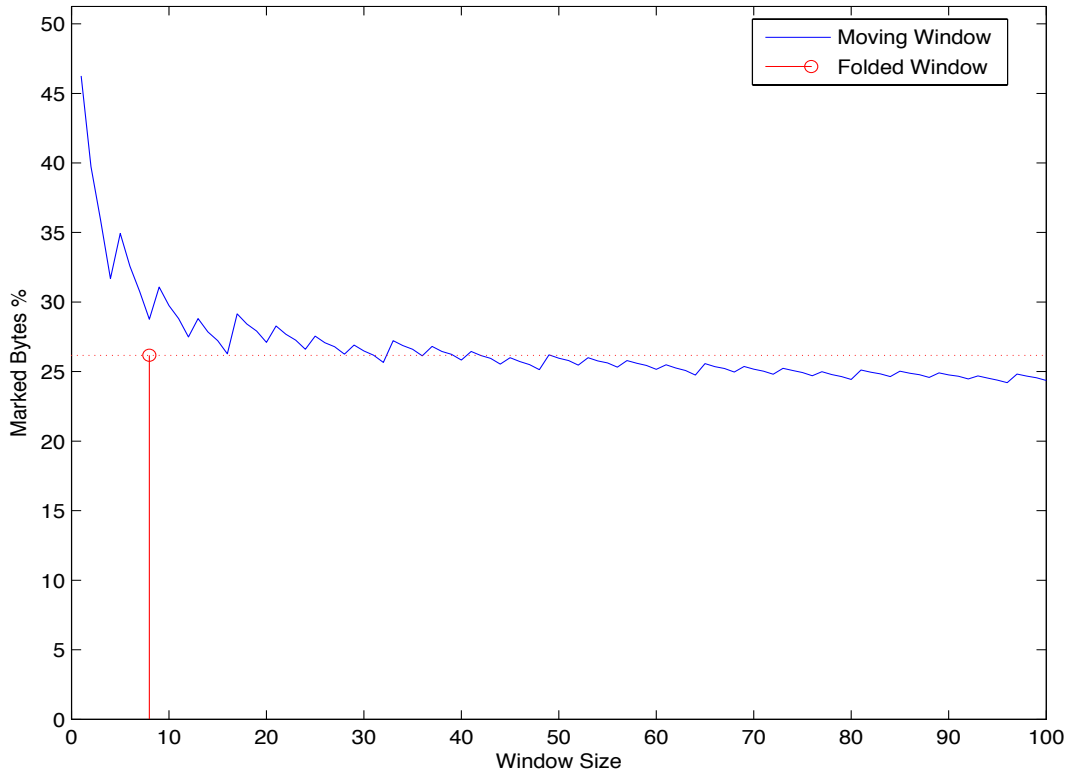


Figure 1. StarWars B3, QP 16. Marked traffic versus the window size, with the use of the FW and the MW mechanism

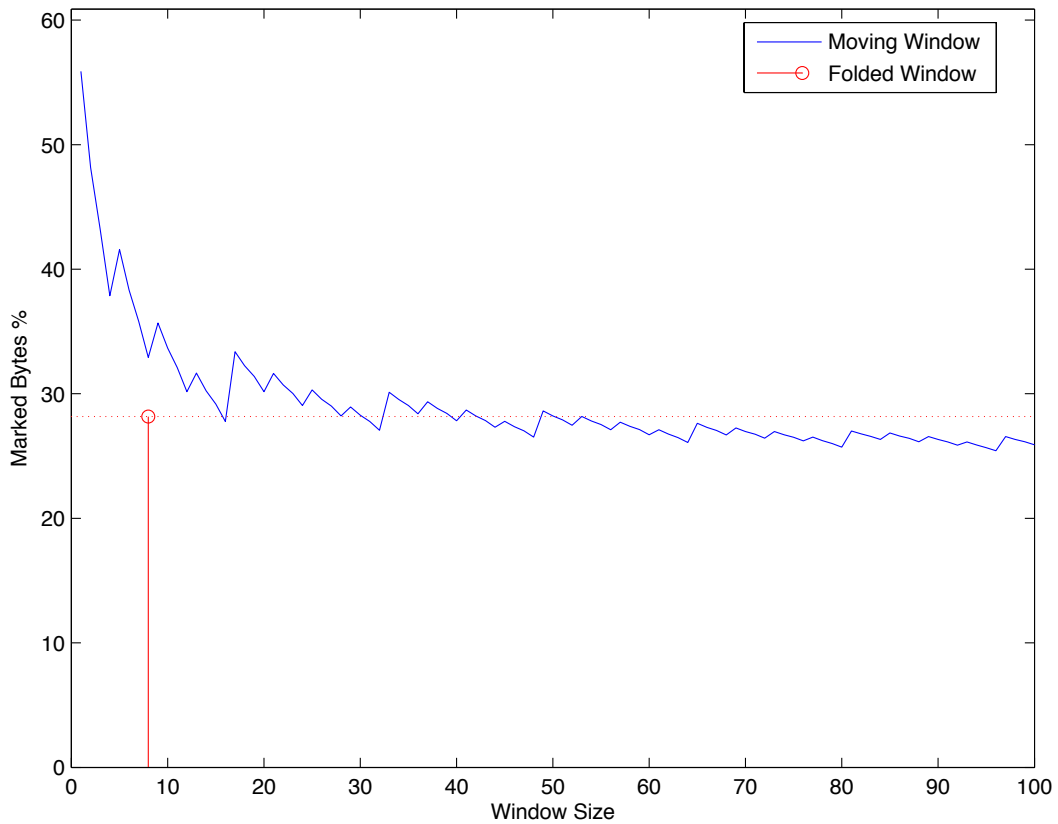


Figure 2. StarWars B3, QP 28. Marked traffic versus the window size, with the use of the FW and the MW mechanism

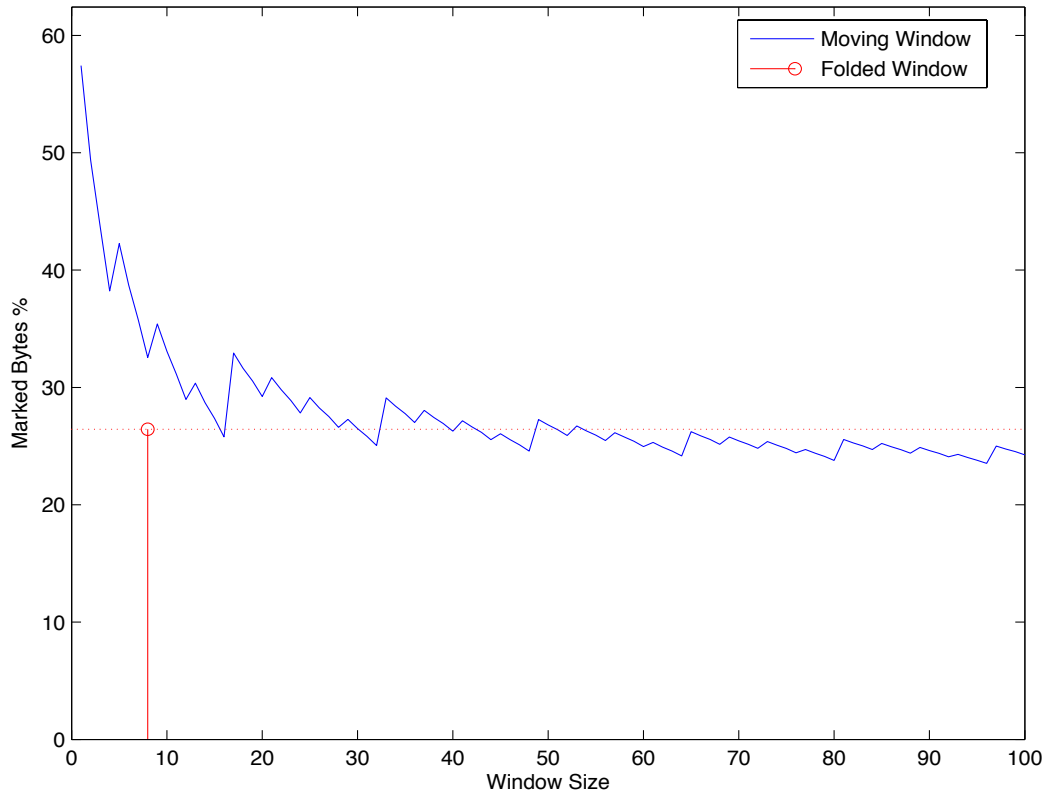


Figure 3 StarWars B3, QP 38. Marked traffic versus the window size, with the use of the FW and the MW mechanism

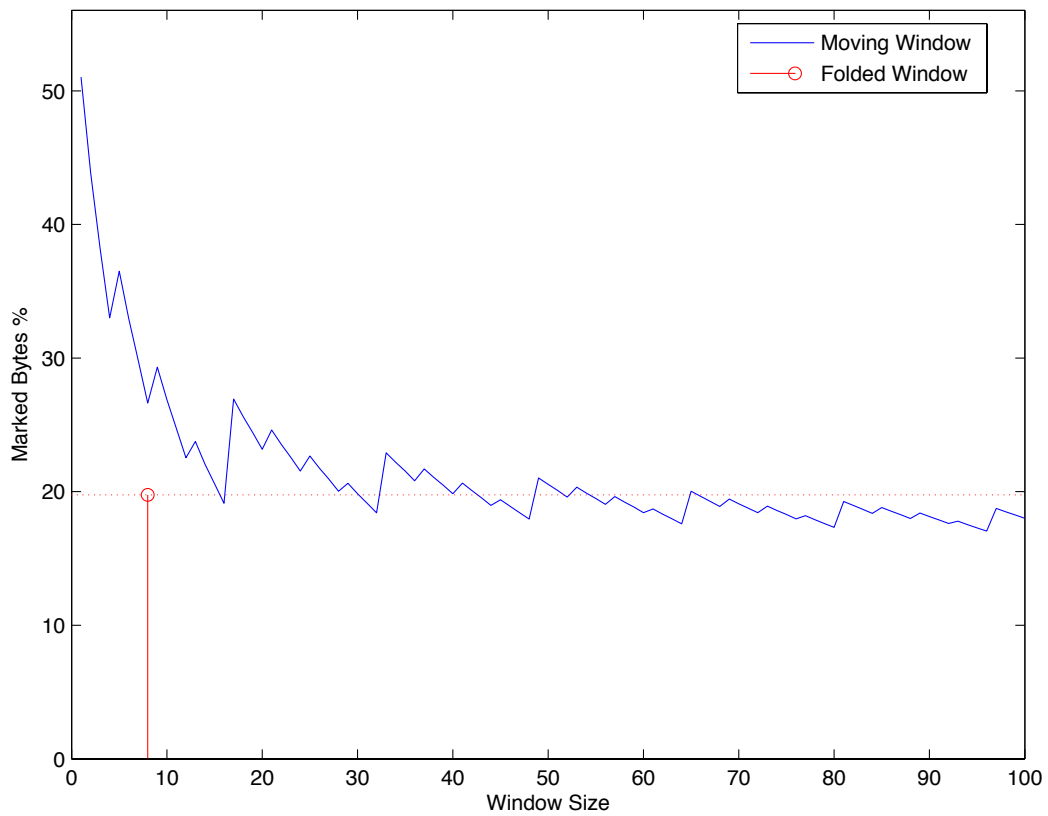


Figure 4 StarWars B3, QP 48. Marked traffic versus the window size, with the use of the FW and the MW mechanism

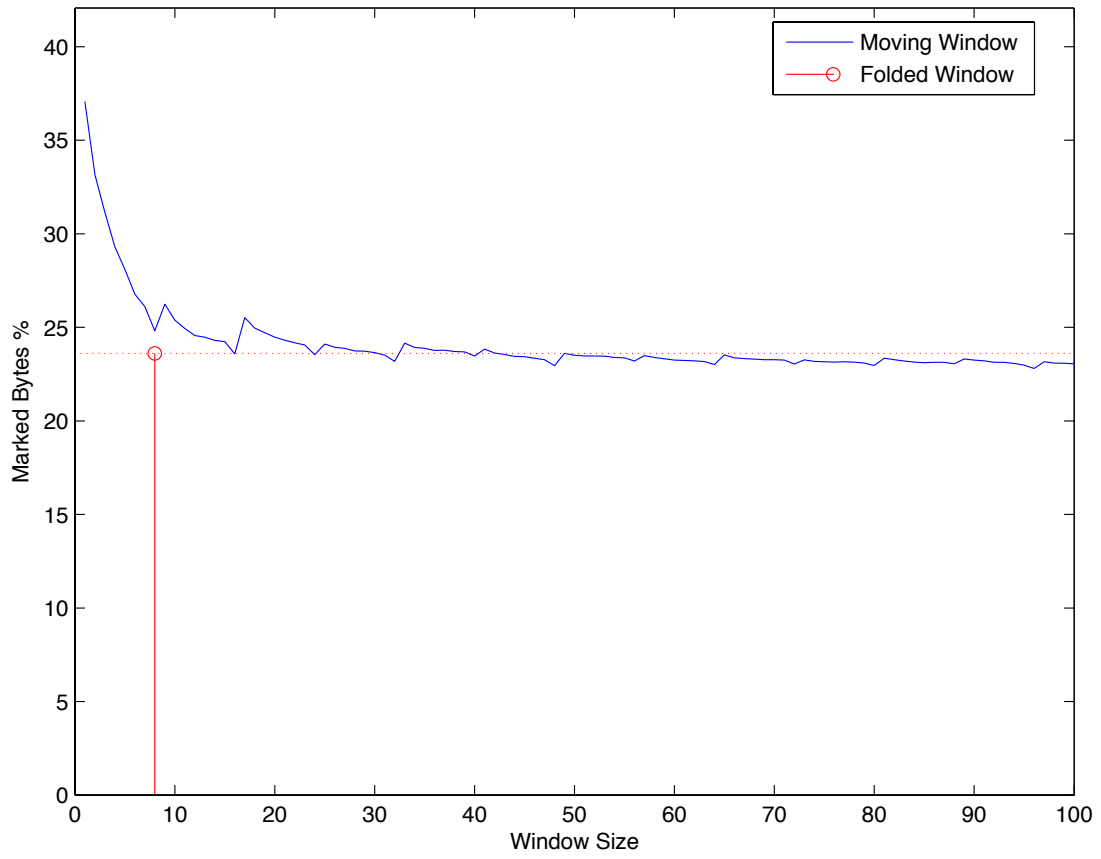


Figure 5 Sonydemo B7, QP 16. Marked traffic versus the window size, with the use of the FW and the MW mechanism

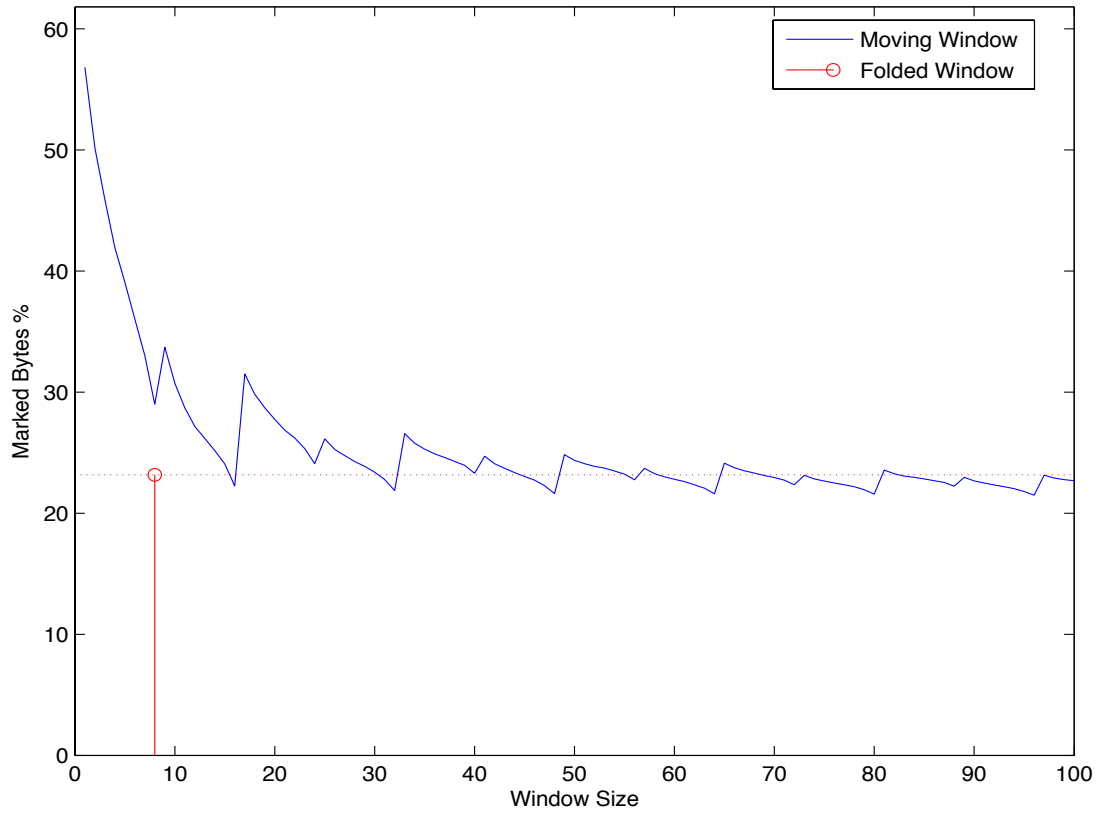


Figure 6 Sonydemo B7, QP 28. Marked traffic versus the window size, with the use of the FW and the MW mechanism

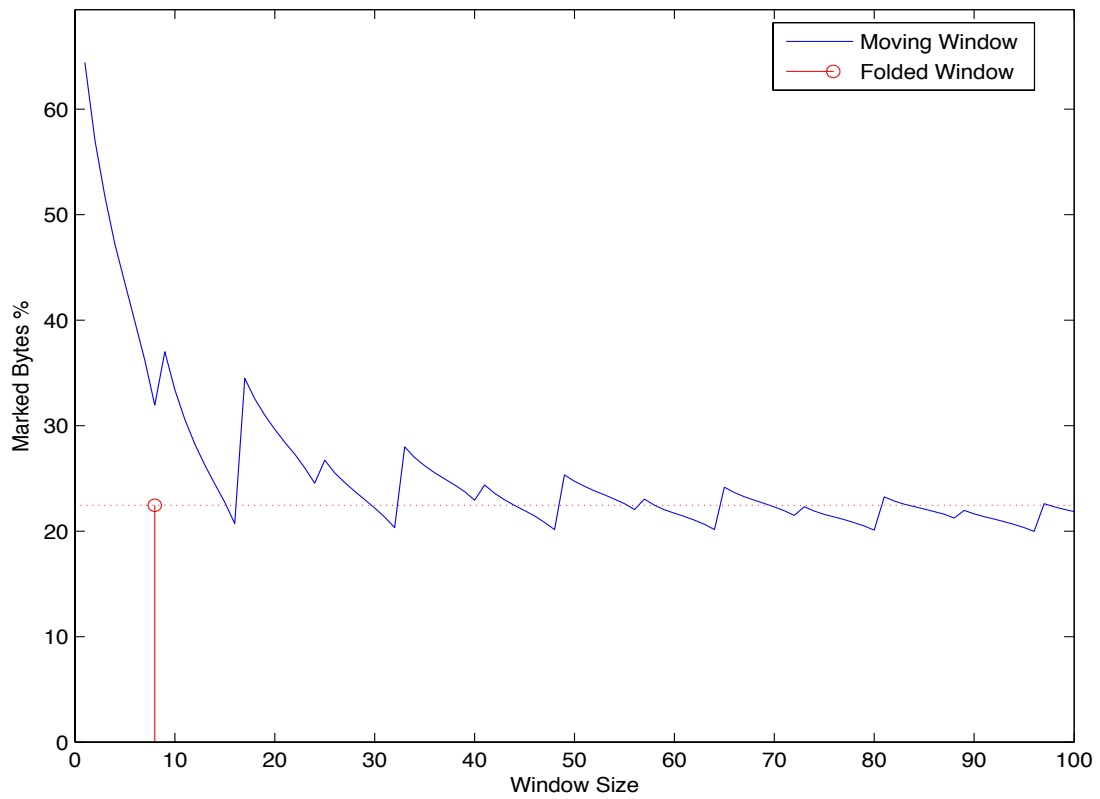


Figure 7 Sonydemo B7, QP 38. Marked traffic versus the window size, with the use of the FW and the MW mechanism

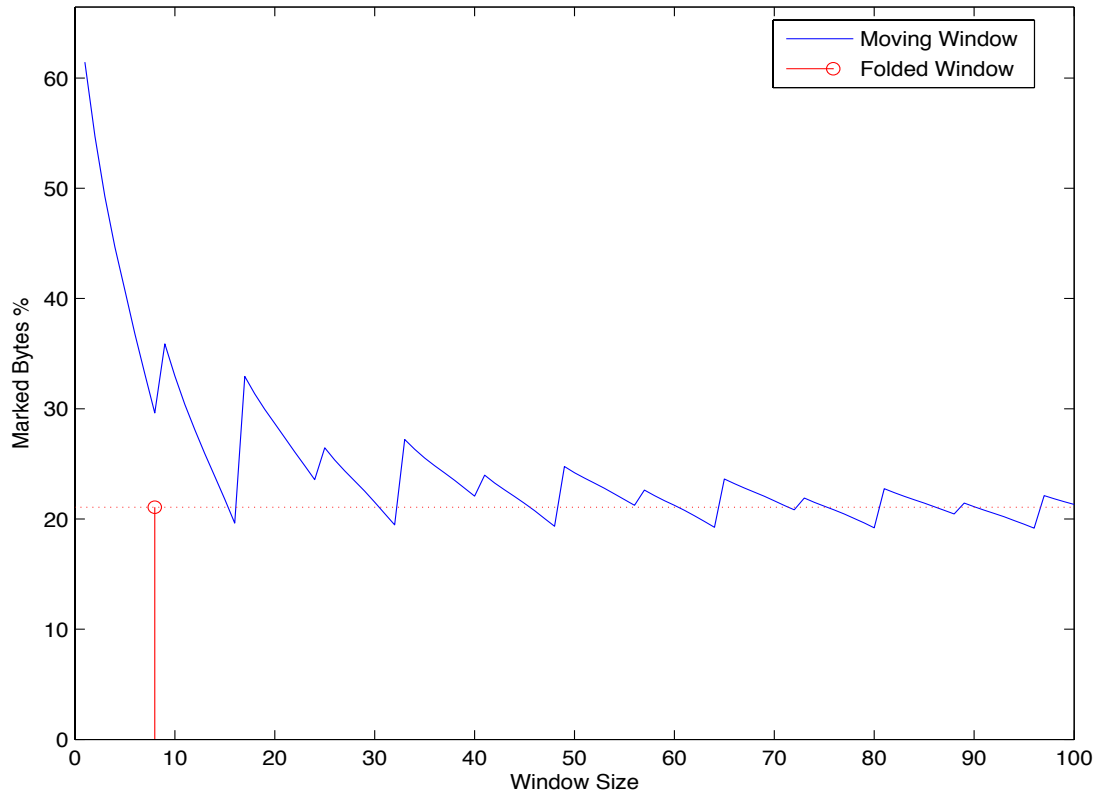


Figure 8 Sonydemo B7, QP 48. Marked traffic versus the window size, with the use of the FW and the MW mechanism

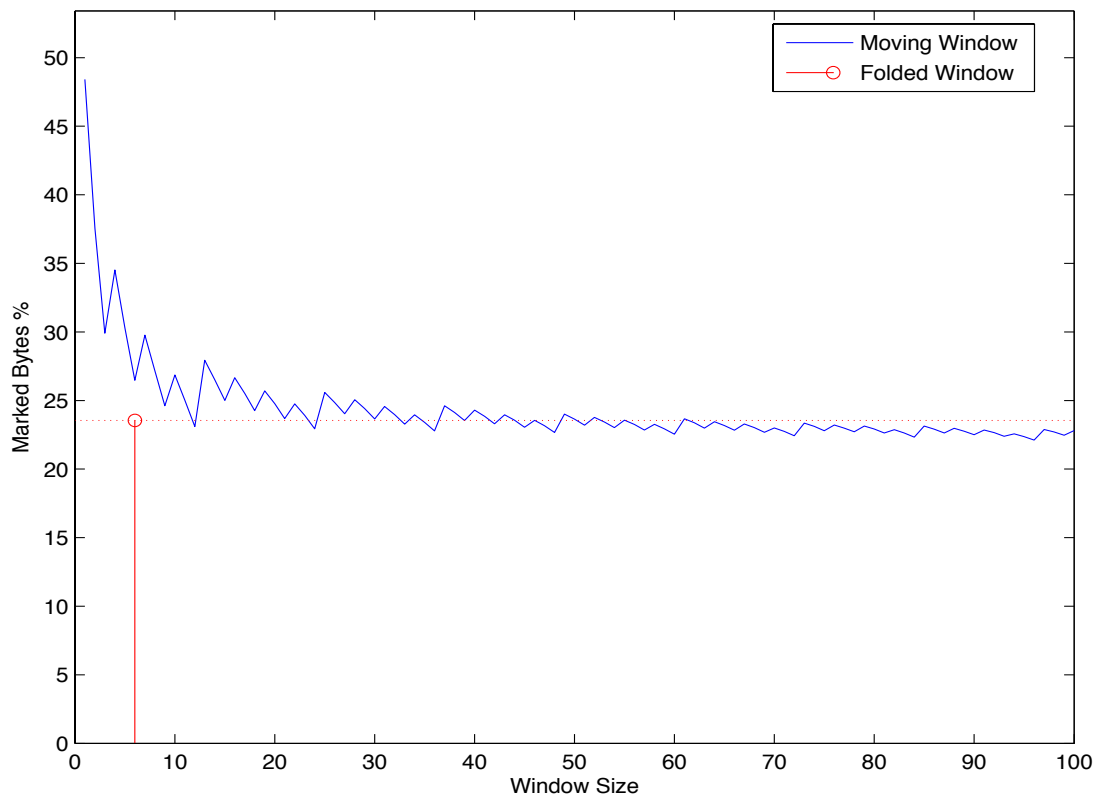


Figure 9 Kaet's from Mars to China B2, QP 28. Marked traffic versus the window size, with the use of the FW and the MW mechanism

Movie	Codec	QP	MW	FW	Comparison	Comparison(%)
Tokyo	B3	16	23.106	22.212	-0.894	-3.9%
		28	31.162	29.669	-1.493	-4.8%
		38	30.033	28.023	-2.010	-6.7%
		48	25.170	22.524	-2.646	-10.5%
	B7	16	23.781	23.108	-0.673	-2.8%
		28	31.688	30.544	-1.144	-3.6%
		38	29.966	28.476	-1.490	-5.0%
		48	25.065	23.272	-1.793	-7.2%
Silence of the Lambs	B3	16	33.759	32.246	-1.513	-4.5%
		28	36.895	33.865	-3.030	-8.2%
		38	33.846	30.367	-3.479	-10.3%
		48	25.385	21.440	-3.945	-15.5%
	B7	16	34.288	33.291	-0.997	-2.9%
		28	36.758	34.700	-2.058	-5.6%
		38	33.163	30.863	-2.300	-6.9%
		48	23.472	20.730	-2.742	-11.7%
Star Wars	B3	16	28.766	26.172	-2.594	-9.0%
		28	32.909	28.159	-4.750	-14.4%
		38	32.543	26.443	-6.100	-18.7%
		48	26.627	19.754	-6.873	-25.8%
	B7	16	28.468	26.795	-1.673	-5.9%
		28	32.155	28.785	-3.370	-10.5%
		38	31.057	26.774	-4.283	-13.8%
		48	25.044	19.897	-5.147	-20.6%
SonyDemo	B3	16	23.983	23.346	-0.637	-2.7%
		28	29.017	22.466	-6.551	-22.6%
		38	33.103	22.415	-10.688	-32.3%
		48	31.562	20.150	-11.412	-36.2%
	B7	16	24.816	23.604	-1.212	-4.9%
		28	29.008	23.162	-5.846	-20.2%
		38	31.956	22.467	-9.489	-29.7%
		48	29.614	21.061	-8.553	-28.9%
Nbc News	B3	16	15.400	14.890	-0.510	-3.3%
		28	26.867	24.075	-2.792	-10.4%
		38	28.059	21.631	-6.428	-22.9%
		48	27.345	18.294	-9.051	-33.1%
	B7	16	15.568	15.137	-0.431	-2.8%
		28	26.852	24.716	-2.136	-8.0%
		38	26.748	21.904	-4.844	-18.1%
		48	25.743	18.720	-7.023	-27.3%
Terminator	B2	28	23.395	22.689	-0.706	-3.0%
		38	22.831	21.598	-1.233	-5.4%
		48	23.530	21.096	-2.434	-10.3%

Kaet's from Mars to China	B2	28	26.478	23.550	-2.928	-11.1%
Kaet's Horizon	B2	28	25.466	12.606	-12.860	-50.5%
SonyDemo2	B2	28	27.074	23.496	-3.578	-13.2%
		38	32.043	22.847	-9.196	-28.7%
		48	33.839	21.730	-12.109	-35.8%

Table 2. Marked traffic (%), with the use of the FW and the MW mechanism, and comparison of the results for window size = GoP/2

The reduction in marked Bytes that is achieved with FW is larger for the B3 than the B7 encoded movies, as shown in Table 3, which presents the averages for each trace over all QPs. The reason is that in a B3 encoding with GoP=16 the number of P frames in a GoP is equal to three, whereas in a B7 encoding there is only one P frame in the GoP. Therefore, in a B7 encoding fourteen out of the sixteen frames of the GoP are B frames, creating thus a "uniformity" of frame sizes and decreasing the effect of the FW mechanism (although FW clearly outperforms MW for both B3 and B7 encodings). On the other hand, in a B7 encoding, the better policing of P frames through the "isolation" of I frames in one half of the window gives improved results. The mean reduction in marked bytes, with the use of the FW, is 17%, over all the traces. Still, we need to point out that, despite the improvement achieved by the FW mechanism, the percentage of marked traffic remains very high, given that we are considering conforming users.

Movie	Codec	MW Mean Marked Traffic	FW Mean Marked Traffic	Comparison	Comparison%
Tokyo	B3	27.4	25.6	-1.8	-6.6%
Tokyo	B7	27.6	26.4	-1.2	-4.3%
Silence of the Lambs	B3	32.5	29.5	-3	-9.2%
Silence of the Lambs	B7	31.9	29.9	-2	-6.3%
Star Wars	B3	30.2	25.1	-5.1	-16.9%
Star Wars	B7	29.2	25.6	-3.6	-12.3%
SonyDemo	B3	29.4	22.1	-7.3	-24.8%
SonyDemo	B7	28.8	22.6	-6.2	-21.5%
Nbc News	B3	24.4	19.7	-4.7	-19.3%
Nbc News	B7	23.7	20.1	-3.6	-15.2%
Terminator	B2	23.3	21.8	-1.5	-6.4%
Kaet's from Mars to China	B2	26.5	23.6	-2.9	-11.1%
Kaet's Horizon	B2	25.5	12.6	-12.9	-50.5%
SonyDemo2	B2	31.0	22.7	-8.3	-26.8%

Table 3. Mean marked traffic (%) for all traces, with the use of the FW and the MW mechanism

4.2 Results with the use of V-EWMA

We derived our results with the use of three different initial values for g (0.2, 0.3, 0.5) and for three window sizes: GoP , $\frac{GoP}{2}$, $\frac{GoP}{4}$. The g parameter increases in steps of 0.05, as shown in the pseudocode in section 3.2.2. The reason we chose this step size is to be on the conservative side: in the case that the algorithm leads to consecutive increases of g , we do not want to allow the video user to transmit large bursts for a significant period of time.

In Figures 10-15, the percentage of marked traffic decreases, as the window size and the initial value of g increase. Both of these results are expected, as a larger window size increases leniency, and an increase in the value of g has the same effect.

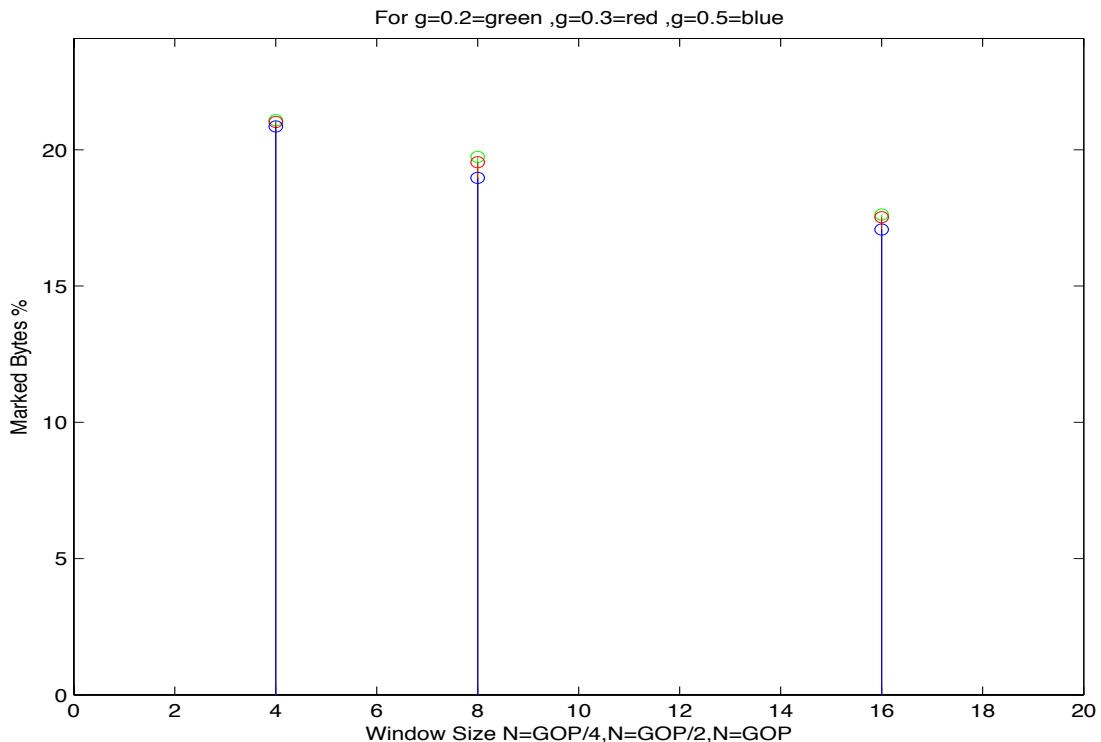


Figure 10 Tokyo B3, QP 16. Marked traffic versus the window size, with the use of the V-EWMA mechanism for three window sizes (4, 8, 16 frames respectively)

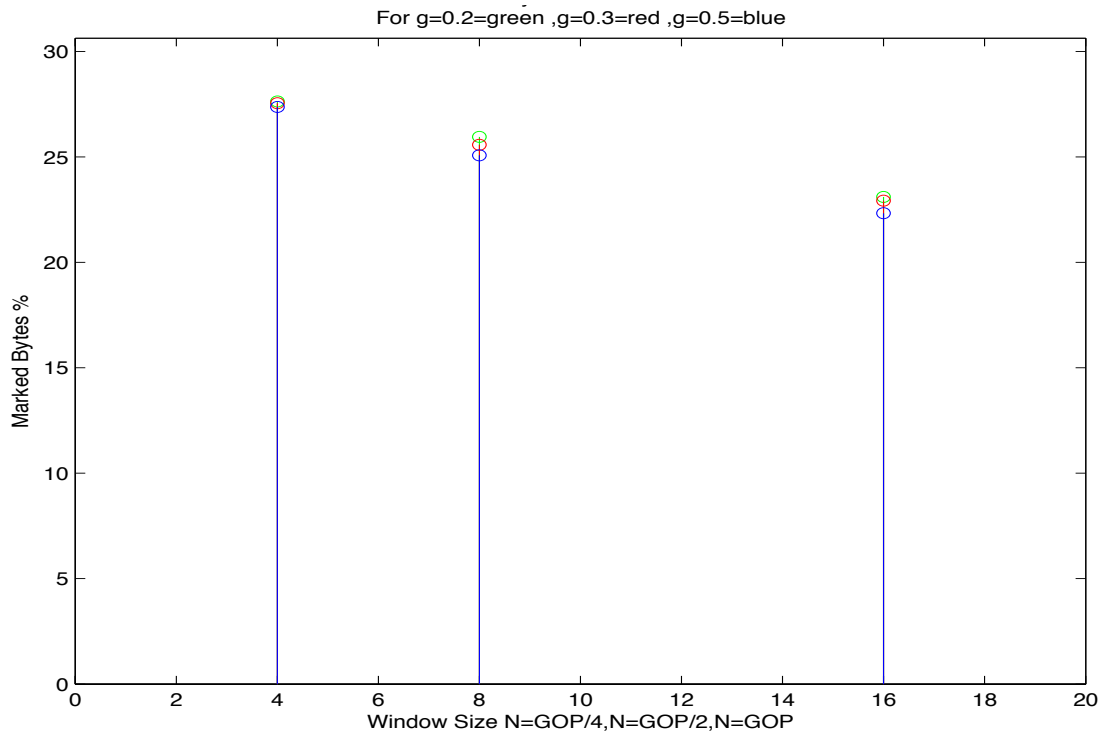


Figure 11 Tokyo B3, QP 28. Marked traffic versus the window size, with the use of the V-EWMA mechanism

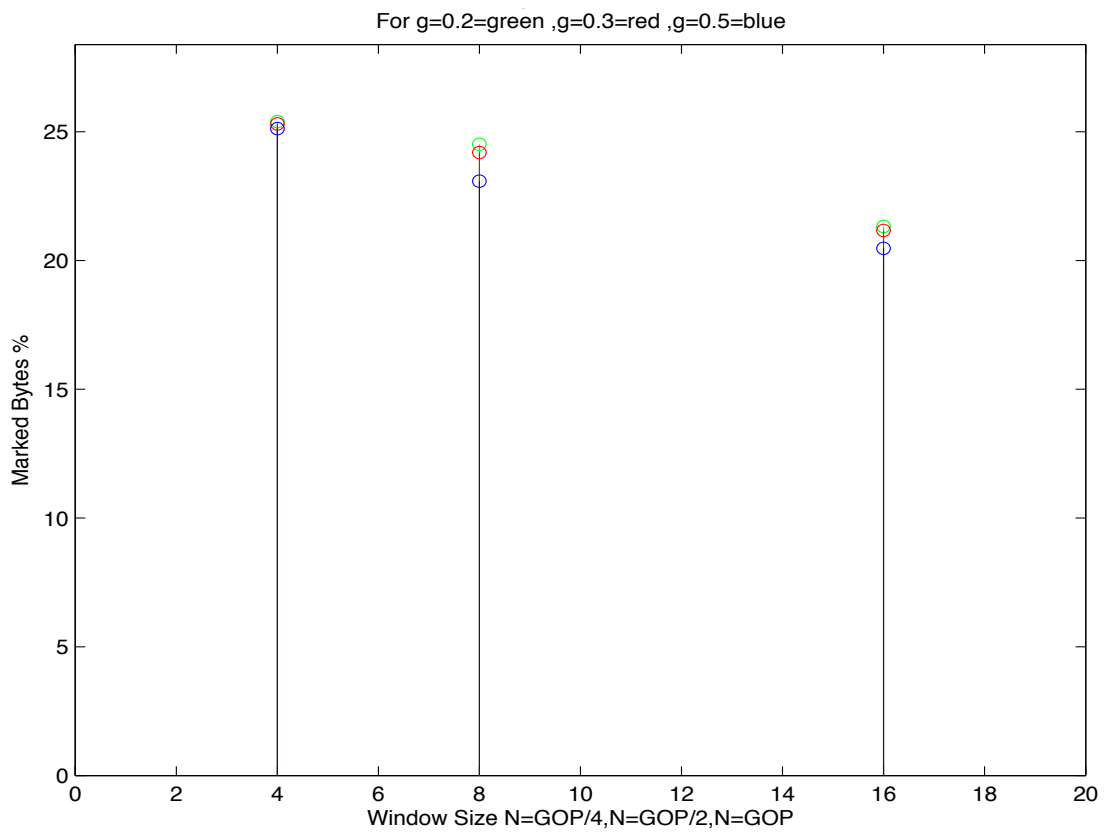


Figure 12 Tokyo B3, QP 38. Marked traffic versus the window size, with the use of the V-EWMA mechanism

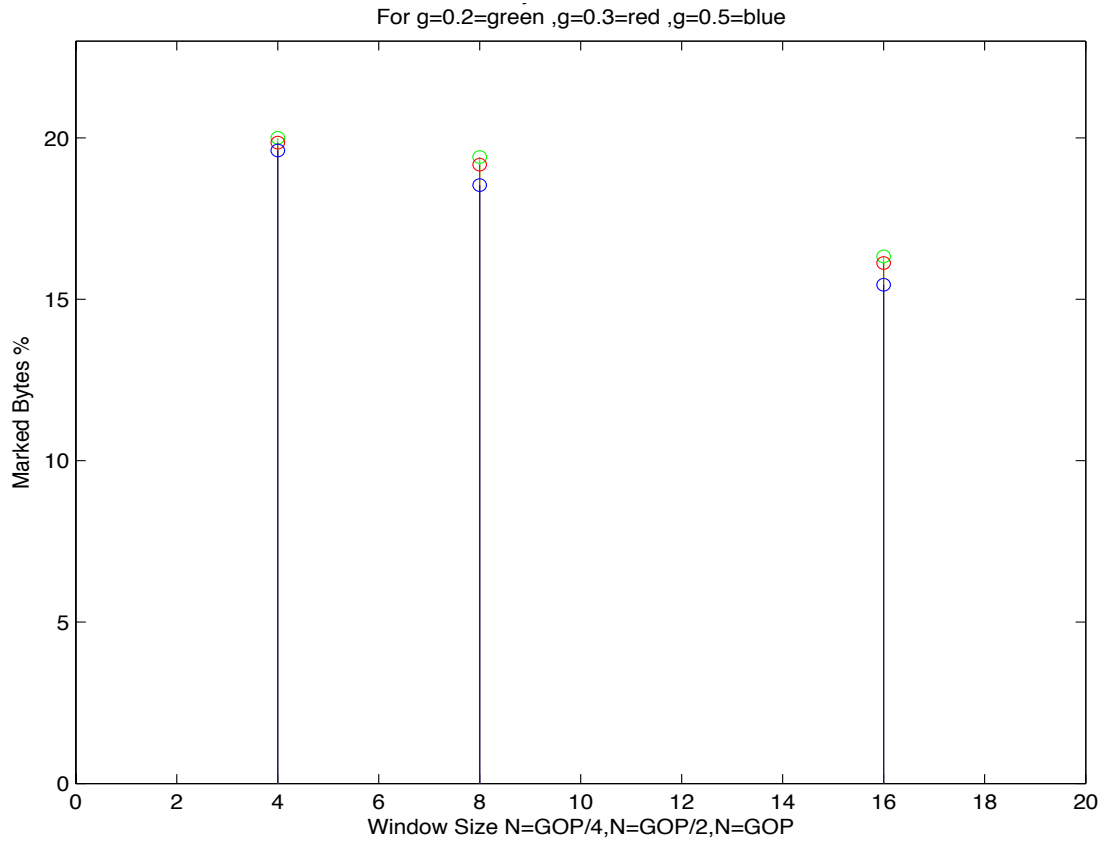


Figure 13 Tokyo B3, QP 48. Marked traffic versus the window size, with the use of the V-EWMA mechanism

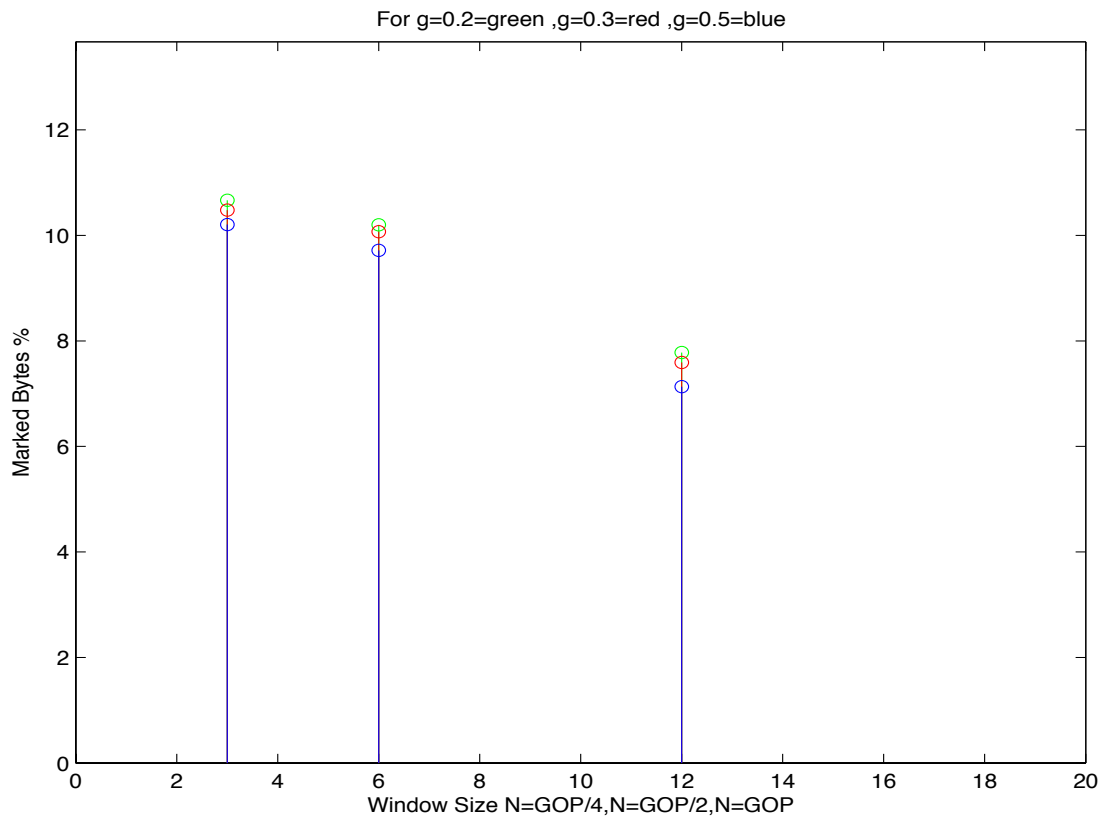


Figure 14 Kaet's horizon B2, QP 28. Marked traffic versus the window size, with the use of the V-EWMA mechanism

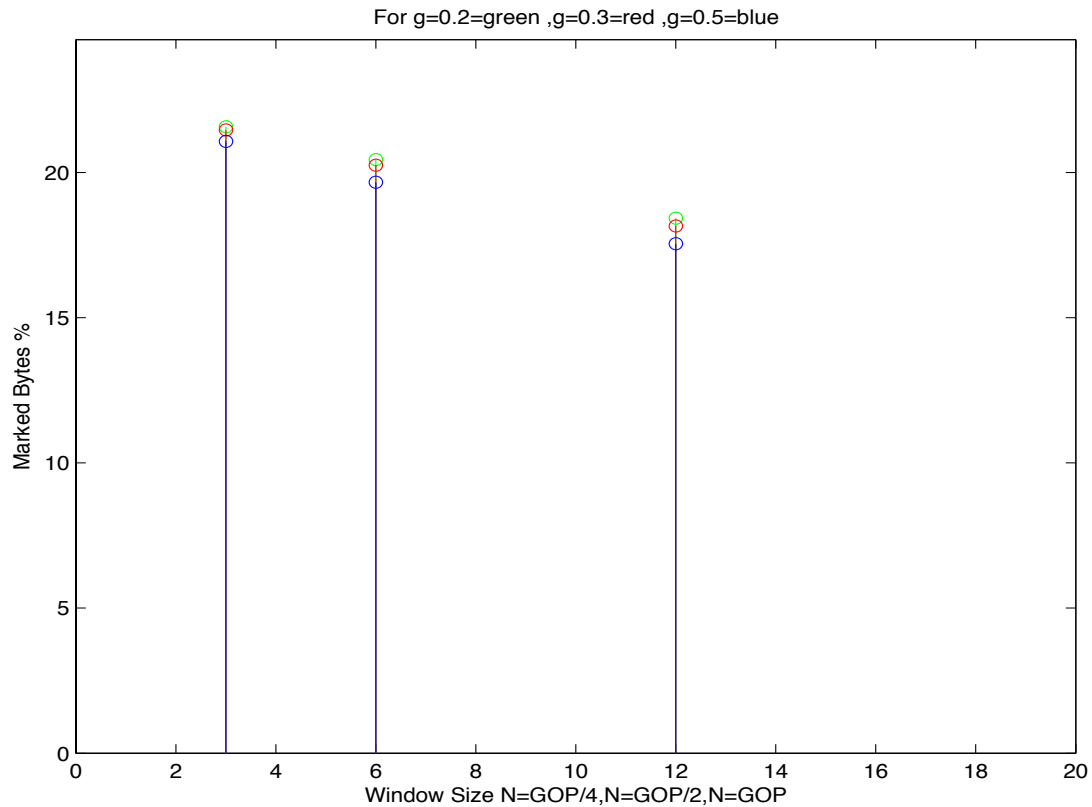


Figure 15 Kaet's from Mars to China B2, QP 28. Marked traffic versus the window size, with the use of the V-EWMA mechanism

Table 4 presents the comparison of the results of EWMA and V-EWMA for conforming users. In these simulations the initial value of g in V-EWMA is set to 0.5, and the constant value of g in EWMA is set to 0.9. The reason for this choice is to make EWMA very lenient, so that we can compare V-EWMA against EWMA while giving EWMA a significant advantage (since we are focusing on conforming users; this choice would backfire if EWMA with $g=0.9$ was implemented for a non-conforming user). For both EWMA and V-EWMA the window size is set equal to GoP.

The results clearly show that V-EWMA excels for all video traces in comparison to EWMA. This is further confirmed by the results presented in Table 5, which presents the average marked traffic for each trace, over all QPs. The results show that the mean marked traffic is larger for the B7 encoded movies than it is for the B3 encoded ones, and that the significant improvement achieved by V-EWMA is comparable, for B3 and B7 encodings. This result is different than the respective result for the FW mechanism (larger reduction for B3 encoded movies). The reason is that the B3 encoding, as already explained, leads to the existence of 3 P frames

within a GoP, whereas the B7 encoded movies contain only 1 P frame in each GoP; the fact that in B3 encodings the frequency of P frames in-between B frames is larger means that it is more difficult to encounter a situation with three consecutive frames each one larger than the next and all of them smaller than the mean. Therefore, g increases more rarely, so the mechanism is less lenient for B3 encoded movies. Still, when g does increase in the policing of B3 encoded movies, this increase allows the system to "breathe" more than it does for B7 encodings, given the larger number of P frames which need to be policed. Hence, the results for B3 and B7 encoded movies are comparable.

Movie	Codec	QP	EWMA	V-EWMA	Comparison	Comparison(%)
Tokyo	B3	16	19.62	17.08	-2.54	-12.96%
		28	25.79	22.33	-3.47	-13.45%
		38	23.73	20.48	-3.26	-13.72%
		48	18.37	15.45	-2.92	-15.90%
	B7	16	20.47	17.63	-2.84	-13.85%
		28	26.61	22.85	-3.76	-14.13%
		38	24.11	20.67	-3.44	-14.26%
		48	19.02	15.82	-3.20	-16.83%
Silence of the Lambs	B3	16	29.21	25.88	-3.33	-11.40%
		28	29.67	27.10	-2.57	-8.65%
		38	25.69	23.43	-2.26	-8.79%
		48	16.95	15.21	-1.74	-10.26%
	B7	16	30.10	27.81	-2.30	-7.63%
		28	30.34	28.00	-2.34	-7.70%
		38	26.18	23.98	-2.20	-8.40%
		48	16.50	14.69	-1.81	-10.94%
Star Wars	B3	16	20.79	16.34	-4.45	-21.41%
		28	21.86	17.29	-4.58	-20.92%
		38	20.23	16.29	-3.94	-19.49%
		48	14.81	12.03	-2.79	-18.82%
	B7	16	21.42	16.57	-4.85	-22.63%
		28	22.63	17.79	-4.84	-21.39%
		38	20.84	16.63	-4.21	-20.22%
		48	15.09	12.26	-2.83	-18.76%
SonyDemo	B3	16	20.35	17.72	-2.63	-12.94%
		28	19.49	17.60	-1.89	-9.68%
		38	18.55	16.47	-2.09	-11.24%
		48	16.96	14.93	-2.03	-11.97%

	B7	16	21.44	18.96	-2.48	-11.56%
		28	20.15	18.43	-1.71	-8.50%
		38	18.77	17.05	-1.72	-9.15%
		48	17.84	15.74	-2.10	-11.77%
Nbc News	B3	16	11.99	10.19	-1.80	-15.04%
		28	17.90	14.81	-3.10	-17.29%
		38	14.69	11.67	-3.02	-20.55%
		48	11.07	8.47	-2.60	-23.45%
	B7	16	12.19	10.27	-1.91	-15.68%
		28	18.51	15.16	-3.35	-18.12%
		38	15.10	11.96	-3.14	-20.79%
		48	11.64	8.87	-2.76	-23.75%
Terminator	B2	28	18.61	15.72	-2.89	-15.55%
		38	17.27	14.50	-2.78	-16.07%
		48	16.39	13.86	-2.53	-15.44%
Kaet's from Mars to China	B2	28	19.91	17.55	-2.36	-11.87%
Kaet's Horizon	B2	28	9.16	7.13	-2.02	-22.12%
SonyDemo2	B2	28	20.70	19.48	-1.21	-5.87%
		38	18.50	17.37	-1.13	-6.12%
		48	18.28	17.03	-1.25	-6.86%

Table 4. Marked traffic (%), with the use of the V-EWMA and the EWMA mechanism

The mean reduction in marked bytes, with the use of V-EWMA, is 15% over all the traces. Our simulations have shown that in order to achieve similar results with V-EWMA, EWMA needs to use $g=0.9$ and a window size at least equal to 60. This, however, would again be a bad choice in the case of policing a non-conforming user, as it will allow the user to violate his contract much more easily.

Finally, we note again, as in Section 4.1 for the FW, that the marked traffic of conforming users remains very high, despite the significant improvement achieved by V-EWMA.

Movie	Codec	EWMA (g=0.9) Window size = GOP Mean Marked Traffic	V-EWMA (g=0.5) Window size = GOP Mean Marked Traffic	Comparison	Comparison%
Tokyo	B3	21.9	18.8	-3.1	-14.2%
Tokyo	B7	22.6	19.2	-3.4	-15.0%
Silence of the Lambs	B3	25.4	22.9	-2.5	-9.8%
Silence of the Lambs	B7	25.8	23.6	-2.2	-8.5%
Star Wars	B3	19.4	15.5	-3.9	-20.1%
Star Wars	B7	20.0	15.8	-4.2	-21.0%
SonyDemo	B3	18.8	16.7	-2.1	-11.2%
SonyDemo	B7	19.5	17.5	-2	-10.3%
Nbc News	B3	13.9	11.3	-2.6	-18.7%
Nbc News	B7	14.4	11.6	-2.8	-19.4%
Terminator	B2	17.2	14.7	-2.5	-14.5%
Kaet's from Mars to China	B2	19.9	17.5	-2.4	-12.1%
Kaet's Horizon	B2	9.2	7.1	-2.1	-22.8%
SonyDemo2	B2	19.2	18.0	-1.2	-6.3%

Table 5. Mean marked traffic (%) for all traces, with the use of the EWMA and the V-EWMA mechanism

4.3 Results with the use of VEWMA-TB

As explained in Section 3.2.3, the large changes in I-frame sizes motivated us to propose this hybrid mechanism. We use the V-EWMA mechanism for I frames and the Token Bucket for policing P and B frames. In this way, we try to "capture" significant changes in activity (given that the changes in I frame size reflect activity changes). This approach also studies how significant the size changes in successive B frames are (since they constitute the vast majority of frames in a GoP). If they are significant, V-EWMA should outperform VEWMA-TB; if not, then the comparison among I frames should show better results for conforming users, hence VEWMA-TB will outperform V-EWMA. By studying the actual sizes of B frames in the respective video files, we had the strong indication that VEWMA-TB would provide better results (size changes were small, for B frames); but we had to implement the algorithm to find out if this was true.

Hence, we compare VEWMA-TB against V-EWMA and EWMA. For V-EWMA and VEWMA-TB the initial g value is set to 0.5, and for EWMA the g value is set to 0.9, as in Section 4.2. The window size is chosen equal to GoP for all mechanisms. Given that the comparison between V-EWMA and EWMA has already been presented in Section 4.2, we only present here the % comparison between our two

proposed mechanisms, V-EWMA and VEWMA-TB (we include EWMA only for completeness of the comparison among the three similar in nature mechanisms).

Our results, presented in Tables 6 and 7, show that VEWMA-TB greatly outperforms V-EWMA (and EWMA, which achieves even worse results).

As shown in Table 7, the reduction in marked traffic is larger for B3 encoded traces than for B7 ones. The reasons for this results are the same with those explained in section 4.1 (more P frames in B3 encoding).

The mean reduction in marked bytes with the use of VEWMA-TB is 38% in comparison to V-EWMA and close to 50% in comparison to EWMA, over all the traces. Our simulations have shown that in order to achieve similar results with VEWMA-TB, EWMA needs to use $g=0.9$ and a window size at least equal to 80.

The excellence of VEWMA-TB in comparison to V-EWMA reveals that the algorithm offers much better QoS to conforming users by solely focusing on the changes of I frames' sizes.

Movie	Codec	QP	EWMA	V-EWMA	VEWMA-TB	Comparison(%) between V-EWMA and VEWMA-TB
Tokyo	B3	16	19.62	17.08	11.85	-30.62%
		28	25.79	22.33	13.16	-41.05%
		38	23.73	20.48	10.50	-48.70%
		48	18.37	15.45	6.97	-54.90%
	B7	16	20.47	17.63	16.80	-4.73%
		28	26.61	22.85	20.12	-11.95%
		38	24.11	20.67	16.57	-19.82%
		48	19.02	15.82	11.49	-27.41%
Silence of the Lambs	B3	16	29.21	25.88	14.79	-42.88%
		28	29.67	27.10	12.21	-54.95%
		38	25.69	23.43	9.33	-60.16%
		48	16.95	15.21	4.91	-67.69%
	B7	16	30.10	27.81	22.08	-20.58%
		28	30.34	28.00	19.15	-31.63%
		38	26.18	23.98	14.56	-39.28%
		48	16.50	14.69	7.23	-50.77%
Star Wars	B3	16	20.79	16.34	12.18	-25.44%
		28	21.86	17.29	10.56	-38.93%
		38	20.23	16.29	8.29	-49.08%
		48	14.81	12.03	4.69	-61.00%

	B7	16	21.42	16.57	15.98	-3.57%
		28	22.63	17.79	16.60	-6.71%
		38	20.84	16.63	13.02	-21.72%
		48	15.09	12.26	7.89	-35.59%
SonyDemo	B3	16	20.35	17.72	12.83	-27.58%
		28	19.49	17.60	8.83	-49.83%
		38	18.55	16.47	6.34	-61.50%
		48	16.96	14.93	4.79	-67.93%
	B7	16	21.44	18.96	17.71	-6.58%
		28	20.15	18.43	14.02	-23.96%
		38	18.77	17.05	10.04	-41.13%
		48	17.84	15.74	8.61	-45.29%
Nbc News	B3	16	11.99	10.19	8.37	-17.87%
		28	17.90	14.81	11.09	-25.10%
		38	14.69	11.67	7.93	-32.03%
		48	11.07	8.47	5.50	-35.03%
	B7	16	12.19	10.27	9.71	-5.52%
		28	18.51	15.16	14.26	-5.96%
		38	15.10	11.96	11.31	-5.43%
		48	11.64	8.87	8.17	-7.91%
Terminator	B2	28	18.61	15.72	10.42	-33.74%
		38	17.27	14.50	8.81	-39.22%
		48	16.39	13.86	7.17	-48.31%
Kaet's from Mars to China	B2	28	19.91	17.55	8.32	-52.61%
Kaet's Horizon	B2	28	9.16	7.13	2.84	-60.21%
SonyDemo2	B2	28	20.70	19.48	8.05	-58.68%
		38	18.50	17.37	6.06	-65.13%
		48	18.28	17.03	5.24	-69.21%

Table 6. Marked traffic (%), with the use of the EWMA, V-EWMA and VEWMA-TB mechanisms.

Movie	Codec	EWMA (g=0.9) Window size = GOP Mean Marked Traffic	V-EWMA (g=0.5) Window size = GOP Mean Marked Traffic	VEWMA-TB (g=0.5) Window size = GOP Mean Marked Traffic	Comparison (%) between V-EWMA and VEWMA-TB
Tokyo	B3	21.9	18.8	10.6	-43.6%
Tokyo	B7	22.6	19.2	16.2	-15.6%
Silence of the Lambs	B3	25.4	22.9	10.3	-55.0%
Silence of the Lambs	B7	25.8	23.6	15.8	-33.1%
Star Wars	B3	19.4	15.5	8.9	-42.6%
Star Wars	B7	20.0	15.8	13.4	-15.2%
SonyDemo	B3	18.8	16.7	8.2	-50.9%
SonyDemo	B7	19.5	17.5	12.6	-28.0%
Nbc News	B3	13.9	11.3	8.2	-27.4%
Nbc News	B7	14.4	11.6	10.9	-6.0%
Terminator	B2	17.2	14.7	8.8	-40.1%
Kaet's from Mars to China	B2	19.9	17.5	8.3	-52.6%
Kaet's Horizon	B2	9.2	7.1	2.8	-60.6%
SonyDemo2	B2	19.2	18.0	6.4	-64.4%

Table 7. Mean marked traffic (%) for all traces, with the use of the EWMA, V-EWMA and VEWMA-TB mechanisms.

4.4 Results with the use of FSA-TB

Since GBTB was shown in [25] to outperform all classic mechanisms, including the original Token Bucket, we compare in this section our proposed policing mechanism, FSA-TB, against GBTB. The results presented in Tables 8 and 9 show that FSA-TB (with a bucket size equal to the peak of the source) not only outperforms GBTB, but also is the only mechanism among all that we have studied and proposed which achieves low percentages of marked traffic (~3% over all the traces). The mean reduction in marked bytes, with the use of the FSA-TB, is 63% over all the traces, in comparison to GBTB.

Movie	Codec	QP	GBTB	FSA-TB	Comparison	Comparison(%)	Coefficient of Variation of Frame Sizes
Tokyo	B3	16	6.9	1.2	-5.7	-82.61%	0.9
		28	10.2	3.4	-6.8	-66.67%	1.6
		38	8.8	4.3	-4.5	-51.14%	1.8
		48	6.7	3.6	-3.1	-46.27%	1.8
	B7	16	7.5	1.1	-6.4	-85.33%	0.8
		28	10.3	3.2	-7.1	-68.93%	1.4
		38	8.2	4.3	-3.9	-47.56%	1.7
		48	7.8	4.0	-3.8	-48.72%	1.8

Silence of the Lambs	B3	16	19.7	3.6	-16.1	-81.73%	1.6
		28	20.5	7.0	-13.5	-65.85%	2.3
		38	17.3	7.9	-9.4	-54.34%	2.5
		48	10.9	5.7	-5.2	-47.71%	2.0
	B7	16	20.1	3.5	-16.6	-82.59%	1.5
		28	20.6	6.9	-13.7	-66.50%	2.3
		38	17.3	8.1	-9.2	-53.18%	2.5
		48	10.9	6.1	-4.8	-44.04%	2.1
Star Wars	B3	16	6.2	0.1	-6.1	-98.39%	1.4
		28	6.7	0.3	-6.4	-95.52%	1.8
		38	7.3	1.2	-6.1	-83.56%	2.0
		48	5.9	1.3	-4.6	-77.97%	1.8
	B7	16	6.5	0.1	-6.4	-98.46%	1.3
		28	6.9	0.2	-6.7	-97.10%	1.8
		38	7.6	1.1	-6.5	-85.53%	2.0
		48	6.4	1.4	-5	-78.13%	1.8
SonyDemo	B3	16	8.6	0.7	-7.9	-91.86%	1.2
		28	11.1	6.0	-5.1	-45.95%	2.2
		38	11.7	7.9	-3.8	-32.48%	2.7
		48	9.4	6.4	-3	-31.91%	2.5
	B7	16	9.2	2.5	-6.7	-72.83%	1.1
		28	10.8	6.0	-4.8	-44.44%	2.1
		38	11.6	8.0	-3.6	-31.03%	2.7
		48	9.6	6.7	-2.9	-30.21%	2.5
Nbc News	B3	16	1.5	0.3	-1.2	-80.00%	0.6
		28	2.2	1.1	-1.1	-50.00%	1.5
		38	2.8	1.4	-1.4	-50.00%	1.9
		48	2.1	1.2	-0.9	-42.86%	2.1
	B7	16	1.5	0.3	-1.2	-80.00%	0.6
		28	2.2	1.1	-1.1	-50.00%	1.4
		38	3.0	1.5	-1.5	-50.00%	1.9
		48	2.3	1.3	-1	-43.48%	2.2
Terminator	B2	28	9.1	0.4	-8.7	-95.60%	1.0
		38	7.2	0.6	-6.6	-91.67%	1.2
		48	5.4	0.9	-4.5	-83.33%	1.5
Kaet's from Mars to China	B2	28	2.2	1.0	-1.2	-54.55%	1.5
Kaet's Horizon	B2	28	2.6	0.7	-1.9	-73.08%	1.8
SonyDemo2	B2	28	10.2	5.8	-4.4	-43.14%	1.8
		38	10.8	7.7	-3.1	-28.70%	2.3
		48	10.1	6.9	-3.2	-31.68%	2.4

Table 8. Marked traffic (%), with the use of the FSA-TB and the GBTB mechanism

Movie	Codec	GBTB Mean Marked Traffic	FSA-TB Mean Marked Traffic	Comparison	Comparison(%)
Tokyo	B3	8.2	3.1	-5.1	-62.2%
Tokyo	B7	8.5	3.2	-5.3	-62.4%
Silence of the Lambs	B3	17.1	6.1	-11	-64.3%
Silence of the Lambs	B7	17.2	6.2	-11	-64.0%
Star Wars	B3	6.5	0.7	-5.8	-89.2%
Star Wars	B7	6.9	0.7	-6.2	-89.9%
SonyDemo	B3	10.2	5.3	-4.9	-48.0%
SonyDemo	B7	10.3	5.8	-4.5	-43.7%
Nbc News	B3	2.2	1.0	-1.2	-54.5%
Nbc News	B7	2.3	1.1	-1.2	-52.2%
Terminator	B2	7.2	0.6	-6.6	-91.7%
Kaet's from Mars to China	B2	2.2	1.0	-1.2	-54.5%
Kaet's Horizon	B2	2.6	0.7	-1.9	-73.1%
SonyDemo2	B2	10.4	6.8	-3.6	-34.6%

Table 9. Mean marked traffic (%) for the movies, grouped by codec, with the use of the GBTB and the FSA-TB mechanism

4.5 Results with the use of GMB-JW

The results for GMB-TB have been derived as the averages of 10 independent runs. Tables 10 and 11 show that even a coarse-grain model, such as the one used in GMB-TB, is enough to provide a substantial decrease in the "unfairly" marked traffic (unfairly in the sense that the users are conforming), in comparison to the classic JW mechanism. However, it is again clear that this decrease is not sufficient - in almost all of the studied cases the percentage of marked traffic was higher than 10%.

The mean reduction in marked bytes, with the use of the GMB-TB, is 18% in comparison to JW, over all the traces.

Movie	Codec	QP	JW	GMB-JW	Comparison	Comparison(%)
Tokyo	B3	16	21.87	19.13	-2.74	-12.52%
		28	28.71	25.01	-3.70	-12.88%
		38	26.67	23.02	-3.65	-13.69%
		48	21.32	17.31	-4.00	-18.78%
	B7	16	22.83	19.88	-2.95	-12.91%
		28	29.69	25.64	-4.05	-13.63%
		38	27.21	23.39	-3.83	-14.06%
		48	22.10	17.35	-4.76	-21.52%

Silence of the Lambs	B3	16	31.55	27.91	-3.64	-11.55%
		28	32.24	27.42	-4.82	-14.96%
		38	28.53	20.94	-7.60	-26.62%
		48	19.91	14.30	-5.61	-28.18%
	B7	16	32.70	29.52	-3.18	-9.74%
		28	33.03	28.08	-4.95	-14.97%
		38	29.10	23.36	-5.74	-19.72%
		48	19.36	15.28	-4.07	-21.04%
Star Wars	B3	16	25.19	22.43	-2.76	-10.96%
		28	26.50	23.24	-3.26	-12.29%
		38	24.59	19.32	-5.27	-21.42%
		48	18.26	14.15	-4.10	-22.48%
	B7	16	26.04	21.99	-4.05	-15.56%
		28	27.43	22.92	-4.51	-16.43%
		38	25.26	20.23	-5.03	-19.91%
		48	18.55	14.11	-4.44	-23.92%
SonyDemo	B3	16	21.87	19.23	-2.64	-12.07%
		28	20.97	17.40	-3.57	-17.01%
		38	19.97	16.53	-3.44	-17.23%
		48	18.17	13.44	-4.73	-26.01%
	B7	16	22.98	20.11	-2.88	-12.51%
		28	21.47	15.90	-5.56	-25.92%
		38	20.02	14.15	-5.88	-29.34%
		48	19.21	13.28	-5.93	-30.84%
Nbc News	B3	16	14.49	12.89	-1.60	-11.02%
		28	22.52	19.96	-2.56	-11.39%
		38	19.29	16.90	-2.39	-12.38%
		48	15.18	12.03	-3.15	-20.73%
	B7	16	14.77	12.72	-2.05	-13.87%
		28	23.33	19.97	-3.36	-14.39%
		38	19.74	16.69	-3.05	-15.45%
		48	15.71	14.03	-1.68	-10.70%
Terminator	B2	28	22.19	19.22	-2.97	-13.38%
		38	20.89	17.70	-3.20	-15.29%
		48	19.92	15.86	-4.06	-20.36%
Kaet's from Mars to China	B2	28	22.61	17.03	-5.58	-24.66%
Kaet's Horizon	B2	28	11.44	8.97	-2.46	-21.53%
SonyDemo2	B2	28	21.88	16.69	-5.19	-23.73%
		38	19.67	14.67	-4.99	-25.39%
		48	19.51	13.67	-5.84	-29.94%

Table 10. Marked traffic (%), with the use of the GMB-JW and the JW mechanism

Movie	Codec	JW Window size = GOP Mean Marked Traffic	GMB-JW Window size = GOP Mean Marked Traffic	Comparison	Comparison%
Tokyo	B3	24.6	21.1	-3.5	-14.2%
Tokyo	B7	25.5	21.6	-3.9	-15.3%
Silence of the Lambs	B3	28.1	22.6	-5.5	-19.6%
Silence of the Lambs	B7	28.5	24.1	-4.4	-15.4%
Star Wars	B3	23.6	19.8	-3.8	-16.1%
Star Wars	B7	24.3	19.8	-4.5	-18.5%
SonyDemo	B3	20.2	16.7	-3.5	-17.3%
SonyDemo	B7	20.9	15.9	-5.0	-23.9%
Nbc News	B3	17.9	15.4	-2.5	-14.0%
Nbc News	B7	18.4	15.9	-2.5	-13.6%
Terminator	B2	21.0	17.6	-3.4	-16.2%
Kaet's from Mars to China	B2	22.6	17.0	-5.6	-24.8%
Kaet's Horizon	B2	11.4	9.0	-2.4	-21.1%
SonyDemo2	B2	20.4	15.0	-5.4	-26.5%

Table 11. Mean marked traffic (%) for the movies, grouped by codec, with the use of the JW and the GMB-JW mechanism

4.6 Comparison of all mechanisms

In this section we compare FSA-TB against all other mechanisms used in our study. We first compare the mechanisms in the case of conforming users, which has been the main subject of this thesis, and then in the case of non-conforming users, which is a top priority for every policing mechanism.

4.6.1 Conforming Users

In this comparison, we have chosen to implement all other mechanisms with parameters which increase their leniency. Of course, this is a bad choice in the case of non-conforming users; it is used here in order to provide a “fairer” comparison for all mechanisms against FSA-TB.

FSA-TB is shown to outperform all mechanisms for every video trace used in our study, and to significantly decrease the percentage of marked traffic on average, as shown in Table 12.

			Average Improvement (Decrease) in Marked Traffic of Conforming Users
FSA-TB	vs.	MW Window Size = GoP/2	90%
FSA-TB		FW Window Size = GoP/2	87%
FSA-TB		JW Window Size = GoP	87%
FSA-TB		EWMA(g=0.9) Window Size = GoP	85%
FSA-TB		GMB-JW Window Size = GoP	83%
FSA-TB		V-EWMA (g=0.5) Window Size = GoP	83%
FSA-TB		TB Bucket Size = 5*Peak	83%
FSA-TB		VEWMA-TB (g=0.5) Window Size = GoP	68%
FSA-TB		GBTB	63%

Table 12. Comparison of FSA-TB with all the other mechanisms used in our study, for conforming users.

Given that GBTB was shown, from all of our results, to be the second best policing mechanism for conforming users, we proceeded to make one more comparison between FSA-TB and GBTB. We have found, via simulation, the required token generation rate for which each mechanism achieves 1% and 0.01% of marked traffic. These rates are presented in Table 13 for each trace used in our study. On average, FSA-TB needs a 15% smaller token generation rate than GBTB to achieve 1% marked traffic, and a 21% smaller token generation rate than GBTB to achieve 0.01% marked traffic over all the traces studied.

An interesting observation is that GBTB is shown in these results to partially “close the gap” between its own performance and that of FSA-TB, in the sense that although FSA-TB is again clearly better, GBTB’s performance is not as far behind as it was in all the previous results, and for some traces GBTB is even marginally better (i.e., it needs a marginally smaller token generation rate to achieve the same low percentage of marked traffic as FSA-TB). The reason for these results can be traced to the nature of both mechanisms. FSA-TB uses the I frames’ sizes to dynamically adjust, on the fly, the token generation rate for P and B frames; i.e., FSA-TB acquires the needed token generation rate for P and B frames on its own, and can only

benefit from a larger token generation rate for I frames. On the other hand, GBTB uses a static token generation rate for each frame type, therefore the increase in token generation rate has a significant positive influence (larger leniency) for all types of frames; hence, a uniform increase of the token generation rate for all frames can significantly improve its performance.

Movie	Codec	QP	Marked Traffic (%) with GBTB	GBTB X times the mean value (for 1%)	GBTB X times the mean value (for 0.01%)	Marked Traffic (%) with FSA-TB	FSA-TB X times the mean value (for 1%)	FSA-TB X times the mean value (for 0.01%)
Tokyo	B3	16	6.9	1.19	1.24	1.2	1.04	1.28
		28	10.2	1.33	1.40	3.4	1.29	1.48
		38	8.8	1.32	1.42	4.3	1.33	1.48
		48	6.7	1.23	1.31	3.6	1.24	1.36
	B7	16	7.5	1.19	1.26	1.1	1.03	1.28
		28	10.3	1.30	1.40	3.2	1.28	1.44
		38	8.2	1.32	1.42	4.3	1.33	1.48
		48	7.8	1.25	1.33	4.0	1.26	1.4
Silence of the Lambs	B3	16	19.7	2.96	3.26	3.6	1.76	2.44
		28	20.5	2.96	3.27	7.0	2.5	3.24
		38	17.3	2.67	3.33	7.9	2.68	3.4
		48	10.9	1.79	3.52	5.7	1.8	2.56
	B7	16	20.1	2.80	3.01	3.5	1.74	2.4
		28	20.6	2.86	3.16	6.9	2.48	3.2
		38	17.3	2.72	3.38	8.1	2.73	3.44
		48	10.9	1.90	2.60	6.1	1.91	2.64
Star Wars	B3	16	6.2	1.48	1.68	0.1	1	1.08
		28	6.7	1.44	1.65	0.3	1	1.24
		38	7.3	1.48	1.58	1.2	1.03	1.52
		48	5.9	1.31	1.56	1.3	1.04	1.6
	B7	16	6.5	1.39	1.55	0.1	1	1.08
		28	6.9	1.42	1.55	0.2	1	1.2
		38	7.6	1.42	1.53	1.1	1.03	1.52
		48	6.4	1.31	1.59	1.4	1.05	1.64
SonyDemo	B3	16	8.6	1.10	1.13	0.7	1	1.16
		28	11.1	1.17	1.23	6.0	1.19	1.28
		38	11.7	1.18	1.22	7.9	1.18	1.28
		48	9.4	1.13	1.16	6.4	1.14	1.2
	B7	16	9.2	1.10	1.18	2.5	1.1	1.24
		28	10.8	1.18	1.23	6.0	1.19	1.28
		38	11.6	1.17	1.21	8.0	1.18	1.28

		48	9.6	1.13	1.16	6.7	1.15	1.2
Nbc News	B3	16	1.5	1.02	1.05	0.3	1	1.12
		28	2.2	1.02	1.77	1.1	1.02	1.16
		38	2.8	1.03	2.98	1.4	1.03	1.2
		48	2.1	1.02	3.97	1.2	1.02	1.24
	B7	16	1.5	1.01	1.06	0.3	1	1.12
		28	2.2	1.02	1.50	1.1	1.02	1.16
		38	3.0	1.03	2.50	1.5	1.03	1.2
		48	2.3	1.02	3.59	1.3	1.02	1.24
Terminator	B2	28	9.1	1.28	1.47	0.4	1	1.2
		38	7.2	1.30	1.71	0.6	1	1.16
		48	5.4	1.29	1.67	0.9	1	1.12
From Mars to China	B2	28	2.2	1.02	1.04	1.0	1	1.08
Horizon	B2	28	2.6	1.91	3.76	0.7	1	1.32
SonyDemo2	B2	28	10.2	1.22	1.28	5.8	1.23	1.32
		38	10.8	1.19	1.24	7.7	1.21	1.28
		48	10.1	1.14	1.18	6.9	1.15	1.24

Table 13. Comparison of FSA-TB with GBTB regarding the required token generation rate to achieve 1% and 0.01% marked traffic for conforming users.

4.6.2 Non-Conforming Users

Table 14 presents the results of all mechanisms used in our study for 4 cases of non-conforming users. Movie 1 is the video that is actually transmitted by malevolent users, who falsely declare in their traffic parameters one of the four other movies shown in the Table.

These results show that the dynamic nature of FSA-TB and the other mechanisms we propose in this work, which exploit knowledge and estimates based on video activity, are sufficient in order to steadily provide comparable (and for FSA-TB in some cases better, i.e., stricter) policing results for non-conforming users when compared against the classic mechanisms.

Movie 1 Transmitted		Movie 2 Declared	Movie 3 Declared	Movie 4 Declared	Movie 5 Declared
Sonydemo B7-QP16		Silence of the Lambs B3-QP16	Star Wars B3-QP16	Tokyo Olympics B3-QP28	Nbc News B7-QP48
Statistics (bytes)		Statistics (bytes)	Statistics (bytes)	Statistics (bytes)	Statistics (bytes)
Mean = 8353 Peak = 65289 Mean_Iframes=33333 Mean_Pframes=15581 Mean_Bframes=6052 Peak_Iframes = 65289 Peak_Bframes=34704 Peak_Pframes=50894 Mean GoP size =126825 Max GoP size = 366825 Stdv_GoP =604260		Mean = 2947 Peak = 51991 Mean_Iframes =12594 Mean_Pframes =5587 Mean_Bframes =1483 Peak_Iframes= 51991 Peak_Bframes = 29348 Peak_Pframes 41195 Mean GoP size =47158 Max GoP size = 439940 Stdv_GoP=370440	Mean = 2976 Peak=32681 Mean_Iframes =13130 Mean_Pframes =5377 Mean_Bframes =1530 Peak_Iframes = 31731 Peak_Bframes=19320 Peak_Pframes=32681 Mean GoP size =47633 Max GoP size = 262418 Stdv_GoP=248948	Mean = 1274 Peak=27850 Mean_Iframes =5474 Mean_Pframes =2425 Mean_Bframes =636 Peak_Iframes = 27850 Peak_Bframes=9512 Peak_Pframes=25762 Mean GoP size =20391 Max GoP size = 131901 Stdv_GoP=126699	Mean = 131 Peak=3721 Mean_Iframes =1056 Mean_Pframes =360 Mean_Bframes =49 Peak_Iframes = 3721 Peak_Bframes=785 Peak_Pframes=3226 Mean GoP size =2107 Max GoP size = 9016 Stdv_GoP=7415
FW	23	73	72	88	98
FSA-TB	2.5	60	58	90	99
VEWMA-TB	18	46	48	55	60
V-EWMA	19	62	60	76	95
GMB-JW	20	33	35	60	97
EWMA	21	68	67	86	98
GBTB	9	74	73	88	98
JW	23	68	67	85	98
TB	21	67	67	85	98
MW	25	68	68	86	98

Table 14. Marked traffic (%) for non-conforming users

5. Conclusions

In this thesis we have proposed, for the first time in the relevant literature to the best of our knowledge, a new paradigm for policing bursty video traffic.

Classic mechanisms such as the Token Bucket, the Moving and the Jumping Window and the Exponentially Weighted Moving Average perform static policing, i.e., policing based on a video source's declared traffic parameters. The five new mechanisms proposed in our work, on the contrary, vary the token generation rate by exploiting knowledge regarding the GoP pattern and by making estimates regarding the expected traffic; these estimates are based on the already transmitted video content (sizes of specific frames and GoPs).

Our mechanisms are shown to provide very significant improvements, in terms of the offered QoS to conforming users, in comparison to the static mechanisms with which they are conceptually similar. This is an important result because the static nature of the classic mechanisms is contradictory with the bursty, dynamic nature of video traffic and this results in very high percentages of marked traffic for conforming users.

Our FSA-TB mechanism, in particular, is shown to not only outperform all other mechanisms (classic and new) but also to provide the best QoS to conforming video users. It is also shown to provide at least comparable and often better policing results in the case of non-conforming users.

Based on all of our results, we believe that this new paradigm we introduce, of activity-based video traffic policing, is very promising for next generation networks in order to handle bursty multimedia traffic.

References

- [1] M. Etoh and T. Yoshimura, "Advances in Wireless Video Delivery," Proceedings of the IEEE, Vol. 93, No. 1, 2005, pp. 111-122.
- [2] S.M. Cherry, "Fiber to the Home," IEEE Spectrum, Vol. 41, No. 1, 2004, pp. 42-43.
- [3] V. Raghunathan, S. Ganeriwal, M. Srivastava, and C. Schurgers, "Energy Efficient Wireless Packet Scheduling and Queuing," ACM Transactions on Embedded Computer Systems, Vol. 3, No. 1, pp. 3-23, 2004.
- [4] G. Procissi, A. Garg, M. Gerla, and M.Y. Sanadidi, "Token Bucket Characterization of Long-Range Dependent Traffic," Computer Communications, Vol. 25, Nos. 11/12, pp. 1009-1017, 2002.
- [5] M. Fiddler and V. Sander, "A Parameter Based Admission Control for Differentiated Services Networks," Computer Networks, Vol. 44, No. 4, pp. 463-479, 2004.
- [6] J.-Y. Le Boudec, "Some Properties of Variable Length Packet Shapers," IEEE/ACM Trans. Networking, Vol. 10, No. 3, pp. 329-337, June 2002.
- [7] J. Sairamesh and N. Shroff, "Limitations and Pitfalls of Leaky Bucket-A Study with Video Traffic," Proc. Third IEEE International Conference on Computer Communications Networks (ICCCN '94), pp. 93-98, 1994.
- [8] N.L.S. Fonseca, G.S. Mayor, and C.A.V. Neto, "On the Equivalent Bandwidth of Self-Similar Sources," ACM Transactions Modeling and Computer Simulation, Vol. 10, no. 2, pp. 104-124, 2000.
- [9] T. Ors and S.P.W. Jones, "Performance Optimizations of ATM Input Control Using an Adaptive Leaky-Bucket Mechanism," Proc. Third IFIP Workshop Performance Modeling Evaluations ATM Networks, 1995.
- [10] E.W. Knightly, "Enforceable Quality of Service Guarantees for Bursty Traffic Streams," Proc. IEEE INFOCOM, pp. 635-642, 1998.
- [11] E.P. Rathgeb, "Modeling and Performance Comparison of Policing Mechanisms for ATM Networks," IEEE Journal on Selected Areas in Communications, Vol. 9, No. 3, pp. 325-334, Apr. 1991.
- [12] P. -Y. Kong, K. -C. Chua, and B. Bensau, "A Novel Scheduling Scheme to Share Dropping Ratio while Guaranteeing a Delay Bound in a MultiCode-CDMA Network," IEEE/ACM Trans. Networking, Vol. 11, No. 6, pp. 994-1006, Dec. 2003.
- [13] C.V.N. Albuquerque, M. Faerman, and O.C.M.B. Duarte, "Implementations of Traffic Control Mechanisms for High Speed Networks," Proc. IEEE Int'l Telecomm. Symp., pp. 177-182, 1998.
- [14] A.R. Reibman and A.W. Berger, "Traffic Descriptors for VBR Video Teleconferencing over ATM Networks," IEEE/ACM Trans. Networking, Vol. 3, No. 3, pp. 329-339, June 1995.
- [15] P. Koutsakis, "Dynamic vs. Static Traffic Policing: A New Approach for Videoconference Traffic over Wireless Cellular Networks", IEEE Transactions on Mobile Computing, Vol. 8, No. 9, 2009, pp. 1153-1166.
- [16] G.R. Ash, Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks, first ed., Morgan Kaufmann, 2006.
- [17] D. Marpe, T. Wiegand and G. Sullivan, "The H.264/MPEG4 Advanced Video Coding Standard and its Applications", IEEE Communications Magazine, vol. 44, no. 8, Aug. 2006, pp. 134-143.
- [18] G. Van der Auwera, P. David and M. Reisslein, "Traffic and Quality Characterization of Single-Layer Video Streams Encoded with the H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension", IEEE Trans. on Broadcasting, Vol. 54, No. 3, September 2008, pp. 698-718.
- [19][Online] <http://trace.eas.asu.edu/h264/index.html>

- [20] M. Dai, Y. Zhang, D. Loguinov, "A Unified Traffic model for MPEG-4 and H.264 video traces", IEEE Transactions on Multimedia, Vol. 11, No. 5, 2009, pp.1010–1023.
- [21] C.-F. Tsai, C.-J. Tsang, F.-C. Ren, and C.-M. Yen, "Adaptive radio resource allocation for downlink OFDMA/SDMA systems," in Proceedings of the IEEE ICC 2007, Glasgow, U.K., pp. 5683–5688.
- [22] D. A. Dyson and Z. J. Haas, "A dynamic packet reservation multiple access scheme for wireless ATM," Mobile Network and Applications (MONET) Journal, Vol. 4, No. 2, pp. 87–99, 1999.
- [23] A. Lazaris and P. Koutsakis, "Modeling Multiplexed Traffic from H.264/AVC Videoconference Streams", Computer Communications Journal, Vol. 33, No. 10, 2010, pp. 1235-1242.
- [24]. [Online]:
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- [25] J. Sotiropoulos, A. Delimargas and P. Koutsakis, "Exploiting GoP Patterns for Video Traffic Control over Wireless Networks", in Proceedings of the IEEE WCNC 2012, Paris, France.
- [26] O. Rose, "Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic", Performance Evaluation Journal, Vol. 30, Nos. 1-2, 1997, pp. 69-85.
- [27] G. Chiruvolu, T. K. Das, R. Sankar, N. Ranganathan, "A Scene-Based Generalized Markov Chain Model for VBR Video Traffic", in Proceedings of the IEEE ICC 1998, Atlanta, USA.
- [28] [Online]
http://www.cisco.com/en/US/tech/tk543/tk545/technologies_tech_note09186a00800a3a25.shtml
- [29]. [Online]: http://blog.nielsen.com/nielsenwire/online_mobile/may-2012-top-u-s-online-video-sites/