Technical University of Crete
Electronic and Computer Engineering Department

Diploma Thesis

Traffic Policing for H.264 Video Traffic Transmission

Delimargas Athanasios

Supervisor: Assistant Professor Polychronis Koutsakis

Chania, July 2011

Acknowledgements

I would sincerely like to thank Assistant Professor Polychronis Koutsakis for his help in every step of this work. His guidance was crucial to its completion. I would also like to thank him for his trust in me and for the truly outstanding cooperation we achieved.

Contents

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. In this work, we initially apply some well-known traffic policing mechanisms to control the transmission of real H.264 video traces. We evaluate the mechanisms' performance and we show that all of them provide unnecessarily strict policing for conforming but bursty H.264 video users. These results lead us to propose a simple and efficient new mechanism, which takes into consideration and exploits the Group-of-Pictures (GoP) pattern of the traces. Our proposed mechanism is shown to outperform the other mechanisms used in our study in terms of providing high QoS to conforming video users, with only a minor tradeoff in terms of its leniency, both for conforming and non-conforming video users.

1. Introduction

Traffic from video services, especially videoconference traffic, is expected to be a substantial portion of the traffic carried by emerging wired and wireless networks [1-2]. The explosive growth of wireless multimedia applications, in particular, 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. Five of the mechanisms which have been most extensively studied (all of them static in nature) are: the Token Bucket, the Leaky Bucket and their variations [3- 11]; 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], we have 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 a new mechanism which takes into account and exploits the Group-of-Picture (GoP) pattern of H.264 video traffic. The mechanism is shown to outperform all the classic mechanisms against which it is compared, in terms of providing high QoS to conforming video users, with only a minor tradeoff in terms of its leniency, both for conforming and 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.

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 briefly describe the five static traffic policing mechanisms which we have used in our study, as well as the new mechanism we propose.

3.1 The Token Bucket (TB)

The token bucket mechanism has been chosen in the recent past as the traffic descriptor for ATM networks and has been widely studied. 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, along with the leaky bucket, the predominant method for network traffic shaping. The two methods have different properties and are used for different purposes. The leaky bucket, which will be discussed in Section 3.2, imposes a hard limit on the source transmission rate, whereas the token bucket allows a certain amount of burstiness (which is necessary for video traffic) while imposing a limit on the average source transmission rate [16].

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:

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

3.2 The Leaky Bucket (LB)

The LB mechanism consists of a counter which is incremented by 1 each time a packet is generated by the source and decremented in fixed intervals as long as the counter value is positive. If the momentary packet arrival rate exceeds the decrementation rate, the counter value starts to increase. It is assumed that the source has exceeded the admissible parameter range if the counter reaches a predefined limit, and suitable actions (e.g., discard or mark packets) are taken on all subsequently generated packets until the counter has fallen below its limit again. In our work, we initially set this limit to be equal to the peak rate of the video trace. We also experimented with the size of the limit, as well as with various leak rates for the bucket, in order to find out for which values the non-comforming traffic fell 1% and 0.01% of the total traffic, respectively.

3.3 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].

3.4 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.

3.5 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

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

which can also be expressed as

$N_i = [N- (1-g)* (g* X_{i-1} +...+g^{i-1} * X_1) - g^{i+1} * So)] / (1-g)$

where S_o 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.

3.6 The GoP-Based Token Bucket (GBTB)

As explained in Section 1, we propose and evaluate a new traffic policing mechanism in this work. The mechanism is especially tailored for video traffic, of any GoP pattern (in our work, it is evaluated over H.264 video traces). By taking into consideration the GoP pattern, the proposed mechanism 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. The motivation behind the proposal of this mechanism will be explained in Section 4.5.

4. Results and Discussion

Our simulations were conducted in a Matlab environment, with an Intel Core 2 duo E600 processor.

4.1. Results with the use of TB

In the initial implementation of the mechanism, we chose the token generation rate to be equal to the mean rate of each trace, and we kept the bucket size equal to the peak of the trace. Subsequently, in order to study the system's behavior when greater lenience is allowed by the policer, we increased the bucket size to 2, 3, 4 and 5 times the peak value, keeping the token generation rate equal to the mean. Finally, we found via simulation the token generation rate which is needed in order to obtain a marked video traffic percentage of less than 1% [21] and 0.01% [22], respectively (in these simulations, the bucket size was again set equal to the peak).

In Table 2 we present our results in terms of the percentage of marked traffic of every movie, when applying the TB mechanism. The first five columns show how the increase (doubling, tripling, etc.) of the bucket size affects our results. It is clear that this increase, in most cases, does not have a very significant impact on the percentage of marked traffic. More specifically, the maximum decrease (for a bucket five times larger than the peak) in marked traffic is 3.32%, on average, for all the formats of the traces. This results offers a significant insight into the nature of the H.264 traces under study. It shows that the video traffic is not so bursty as to need the increase in the bucket size, i.e., the cases where the source transmits for a while at a smaller rate than the mean and then transmits consecutive very large frames, close to the peak, are quite rare, if existing at all. On the contrary, the sources often transmit above the mean but not close to the peak, and as a result the percentage of marked traffic is very high, if we take into account that we are studying the case of *conforming users*, i.e., users who, in the long run, do not violate their traffic contract with the provider.

Therefore, we reach the conclusion that the sources actually need more leniency in terms of the token generation rate, instead of a larger bucket size. It is clear, from the last two columns, that the required token generation rate in order to achieve a small percentage of marked traffic varies widely, depending on the burstiness of each trace. In the case of the "Silence of the Lambs" trace, a token generation rate equal to almost 15 times larger than the mean rate is needed for some formats of the trace in order to achieve a 0.01% percentage of marked traffic, while for the much less bursty NBC News trace the token generation rate needs to be 2-4 times larger than the mean rate.

Movie	**Codec**	**QP**	**Marked Traffic for Bucket=Peak**	**2 x Bucket**	**3 x Bucket**	**4 x Bucket**	**5 x Bucket**	**Token Generation Rate, X times the mean value (for 1% marked traffic)**	**Token Generation Rate, X times the mean value (for 0.01% marked traffic)**
Tokyo	B3	16	21.1	20.4	19.8	19.3	18.9	2.4	3.5
		28	26.9	25.5	24.6	23.9	23.3	3.4	5.7
		38	24.1	22.6	21.7	20.9	20.2	4.1	6.2
		48	19.1	17.7	16.8	16.1	15.5	2.3	3.9
	B7	16	22.1	21.3	20.7	20.2	19.8	2.5	3.6
		28	27.8	26.5	25.6	24.9	24.3	3.6	5.7
		38	24.5	23.1	22.1	21.3	20.6	3.0	5.9
		48	19.5	18.1	17.1	16.5	15.9	2.4	4.0
Silence of the Lambs	B3	16	30.1	29.1	28.5	28.0	27.5	6.1	9.3
		28	29.6	28.0	27.2	26.7	26.4	8.9	14.5
		38	25.0	23.6	22.7	22.0	21.6	7.0	14.4
		48	16.6	15.5	14.8	14.1	13.7	2.7	7.5
	B7	16	31.2	30.1	29.4	28.8	28.3	6.3	9.2
		28	30.3	28.7	27.9	27.4	27.0	9.4	14.6
		38	25.5	24.0	23.1	22.4	21.9	7.4	14.6
		48	16.0	14.9	14.2	13.7	13.3	2.8	7.7
Star Wars	B3	16	23.7	21.9	20.8	19.9	19.3	2.7	5.6
		28	24.2	22.2	21.0	20.0	19.2	3.2	8.4

		38	21.4	19.4	18.2	17.5	16.9	2.9	9.3
		48	14.7	13.4	12.6	12.0	11.4	2.1	7.0
	B7	16	24.5	22.8	21.7	20.8	20.2	2.7	5.7
		28	25.1	23.2	21.9	20.9	20.1	3.3	8.7
		38	22.1	20.1	18.8	18.0	17.4	3.1	10.5
		48	14.9	13.5	12.7	12.0	11.4	2.2	6.8
SonyDemo	B3	16	21.4	20.9	20.4	20.0	19.6	1.9	2.7
		28	19.9	19.1	18.5	18.0	17.6	2.0	3.4
		38	18.5	17.3	16.6	16.0	15.4	1.9	3.4
		48	17.0	16.1	15.4	14.7	14.3	1.7	2.3
	B7	16	22.5	22.0	21.6	21.2	20.9	1.9	2.6
		28	20.6	19.8	19.3	18.8	18.4	2.0	2.7
		38	18.7	17.6	16.9	16.4	15.8	1.9	2.5
		48	17.8	16.9	16.1	15.4	14.8	1.8	2.7
Nbc News	B3	16	13.9	13.1	12.6	12.2	11.9	1.7	2.3
		28	20.6	18.7	17.7	17.1	16.6	2.6	3.9
		38	16.0	14.1	13.1	12.4	11.9	2.3	4.6
		48	11.4	9.7	8.7	8.0	7.6	1.8	3.7
	B7	16	14.2	13.3	12.8	12.4	12.0	1.7	2.3
		28	21.2	19.3	18.3	17.6	17.1	2.6	3.9
		38	16.4	14.5	13.4	12.8	12.3	2.5	4.7
		48	11.8	10.1	9.0	8.4	7.9	1.9	3.9
Terminator		28	20.7	19.5	18.6	17.8	17.1	2.1	3.7
		38	19.3	17.9	16.9	16.0	15.1	2.1	2.9
		48	17.8	16.3	15.3	14.4	13.6	2.4	3.7
Mars to China			20.8	19.4	18.5	17.8	17.2	2.2	4.0
Horizon			9.4	8.2	7.3	6.6	6.0	2.1	4.3
SonyDemo2		28	21.1	20.4	19.9	19.4	18.9	2.0	3.1
		38	18.4	17.5	17.0	16.7	16.4	1.8	3.8
		48	18.3	17.2	16.5	16.1	15.7	1.9	3.4

Table 2. Marked traffic (%) for various bucket sizes and token generation rates, with the use of the Token Bucket mechanism.

4.2. Results with the use of LB

Similarly to our approach for the TB mechanism, in our evaluation of the LB mechanism we set the leak rate of the bucket equal to the mean rate of the trace. However, we imposed a harder policing scheme by using a bucket size equivalent to 70%, 80% and 90%, respectively, of the peak rate of each trace.

Our results are shown in Table 3. As expected, for smaller bucket sizes the percentage of the marked traffic is larger and it varies considerably for each trace and for each format of the traces. It is also larger than the percentage of marked traffic with the use of the TB. Three important conclusions can be derived from the results presented in the Table, together with the results shown in Table 2:

a. The increase in bucket size results, in most cases, in a minimal improvement of the percentage of the marked traffic. This confirms the nature of our results for the TB mechanism, i.e., the bucket size is not the crucial parameter of the policing mechanism, for the H.264 traces under study.

b. The percentage of the marked traffic is very high, for all the traces. This fact, together with our previous conclusion, calls for a different leak rate, in order to improve video QoS.

c. By studying the results of both Table 2 and Table 3, we observe that, for the same trace, marked traffic is higher when more successive B frames are present in the GoP (B7). This shows (and it is confirmed from our work in H.264 video traffic modeling [23]) that the burstiness of B frames' sizes is quite high.

Movie	Codec	QP	LB with Bucket Size 70% of the peak	LB with Bucket Size 80% of the peak	LB with Bucket Size 90% of the peak
Tokyo	B3	16	21.4	21.3	21.2
		28	27.4	27.2	27.0
		38	24.7	24.5	24.3
		48	19.7	19.5	19.3
	B7	16	22.4	22.3	22.2
		28	28.4	28.2	28.0
		38	25.2	24.9	24.7
		48	20.2	20.0	19.7
Silence of the Lambs	B3	16	30.6	30.4	30.2
		28	30.2	30.0	29.8
		38	25.8	25.5	25.3
		48	17.2	17.0	16.8
	B7	16	31.7	31.5	31.4
		28	30.9	30.7	30.5
		38	26.3	26.0	25.7
		48	16.6	16.4	16.2
Star Wars	B3	16	24.6	24.2	23.9
		28	25.3	24.9	24.5
		38	22.4	22.0	21.7
		48	15.4	15.2	14.9
	B7	16	25.4	25.0	24.7
		28	26.2	25.8	25.4
		38	23.1	22.7	22.4
		48	15.6	15.3	15.1
SonyDemo	B3	16	21.9	21.6	21.5
		28	20.3	20.1	20.0
		38	18.9	18.8	18.6
		48	17.4	17.2	17.1
	B7	16	23.1	22.8	22.7
		28	21.0	20.8	20.7
		38	19.2	19.0	18.9
		48	18.3	18.1	18.0
Nbc News	B3	16	14.4	14.2	14.1
		28	21.7	21.2	20.8
		38	17.1	16.7	16.3
		48	12.3	11.9	11.6
	B7	16	14.6	14.4	14.3

	28	22.5	22.0	21.6
	38	17.5	17.1	16.7
	48	12.7	12.3	12.1
Terminator	28	21.2	21.0	20.8
	38	19.8	19.6	19.4
	48	18.5	18.2	18.0
From Mars to China		21.4	21.2	21.0
Horizon		10.0	9.8	9.6
SonyDemo2	28	21.3	21.2	21.1
	38	18.8	18.7	18.5
	48	18.7	18.6	18.5

Table 3. Marked traffic (%) for various bucket sizes, with the use of the Leaky Bucket mechanism.

4.3. Results with the use of JW and MW

We applied both of these mechanisms on all the traces used in our study, and with various degrees of strictness in terms of their implementation. 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. As intuitively expected, and shown in Figures 1-32 the percentage of marked traffic decreases as the window size increases, for all the traces. We present a large number of our results here, but not of all them, in order to avoid repetition, since the rest of our results are similar in nature with the ones shown in the Figures.

Three important conclusions which can be derived by all the Figures are that:

a. for window sizes larger than the GoP, the decrease in the percentage of marked traffic is small or negligible, depending on the trace. This seems to indicate that any significant bursts are most often limited within the duration of one GoP, and smoothed out for longer time intervals, for all the traces under study. This behavior of the H.264 traces will be further studied in a diploma thesis in progress, conducted by our research group.

b. As expected by the nature of the MW mechanism, its results exhibit fluctuation, as the window size increases. This is due to the increased "memory" of the MW mechanism, in which each video frame is used by the policer for T consecutive windows, where T is the window size.

c. Again, as intuitively expected by the nature of both mechanisms, the results of the MW mechanism are stricter, for all traces, than those of the JW. The larger "memory" of MW is responsible for this difference, which leads on average to a marked traffic percentage around 1.1 times higher than the one resulting from the use of the JW.

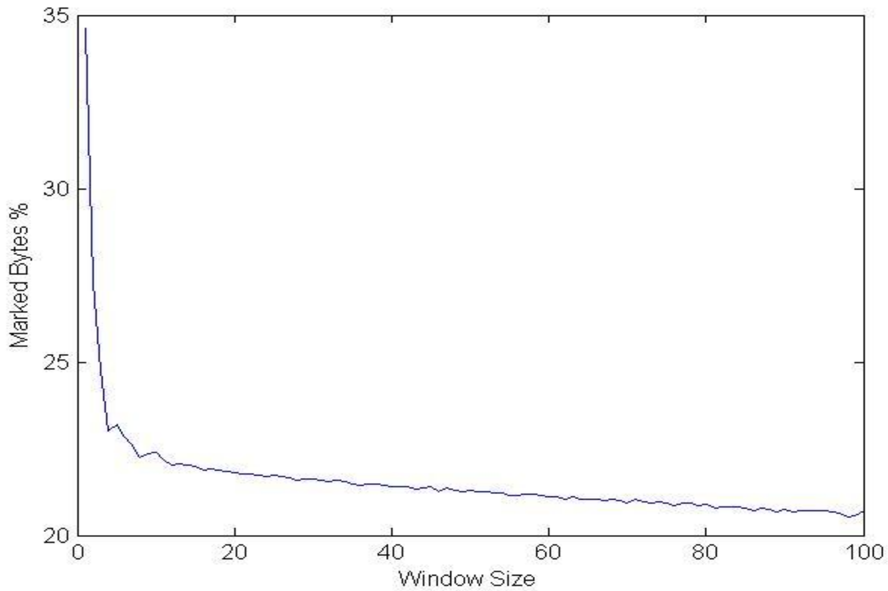

Figure 1. Tokyo Olympics B3, QP 16. Marked traffic versus the window size, with the use of the JW mechanism.

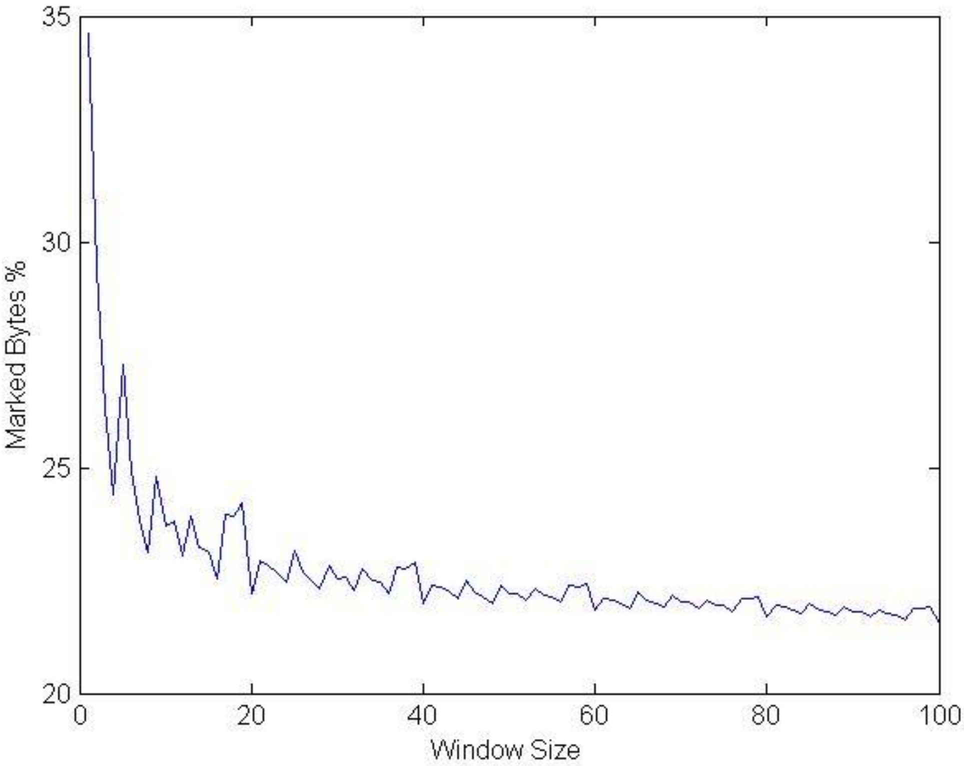

Figure 2. Tokyo Olympics B3, QP 16. Marked traffic versus the window size, with the use of the MW mechanism.

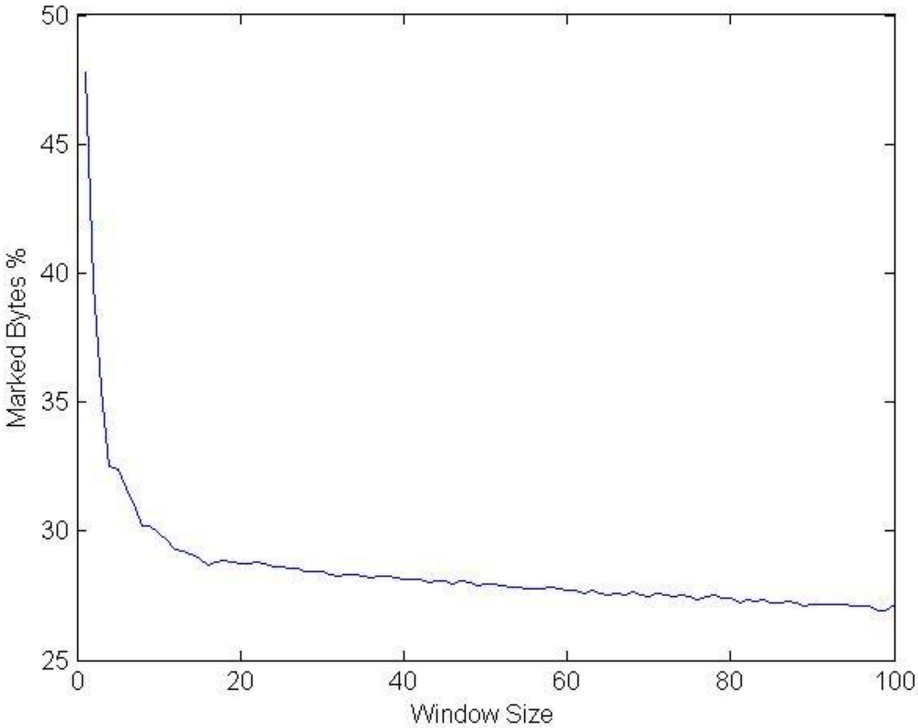

Figure 3. Tokyo Olympics B3, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

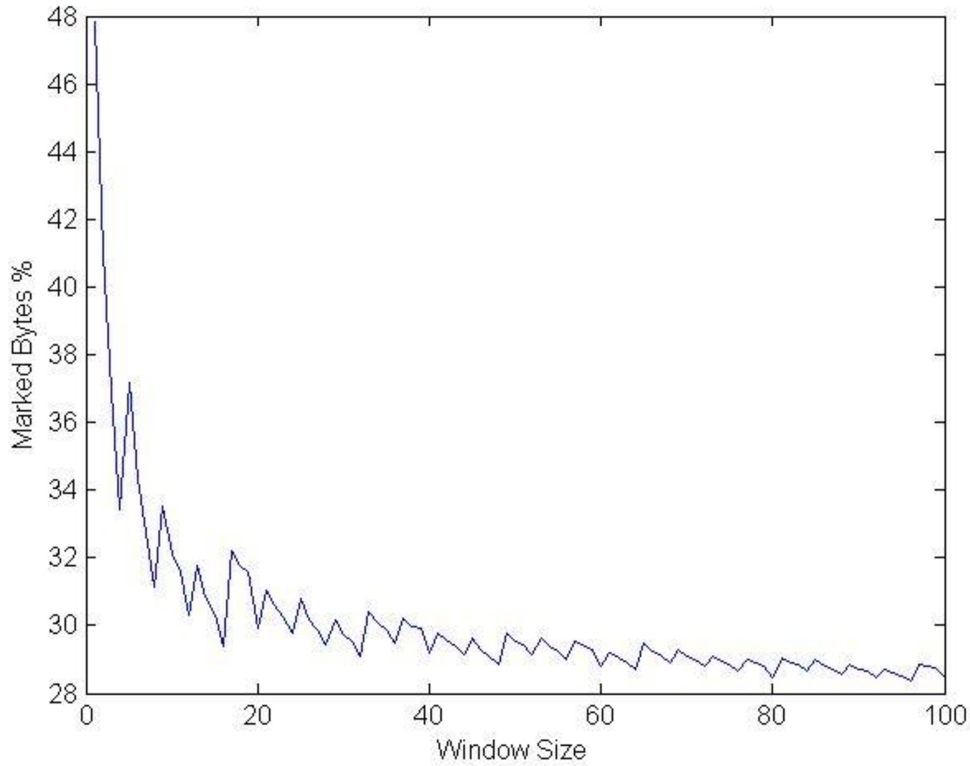

Figure 4. Tokyo Olympics B3, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

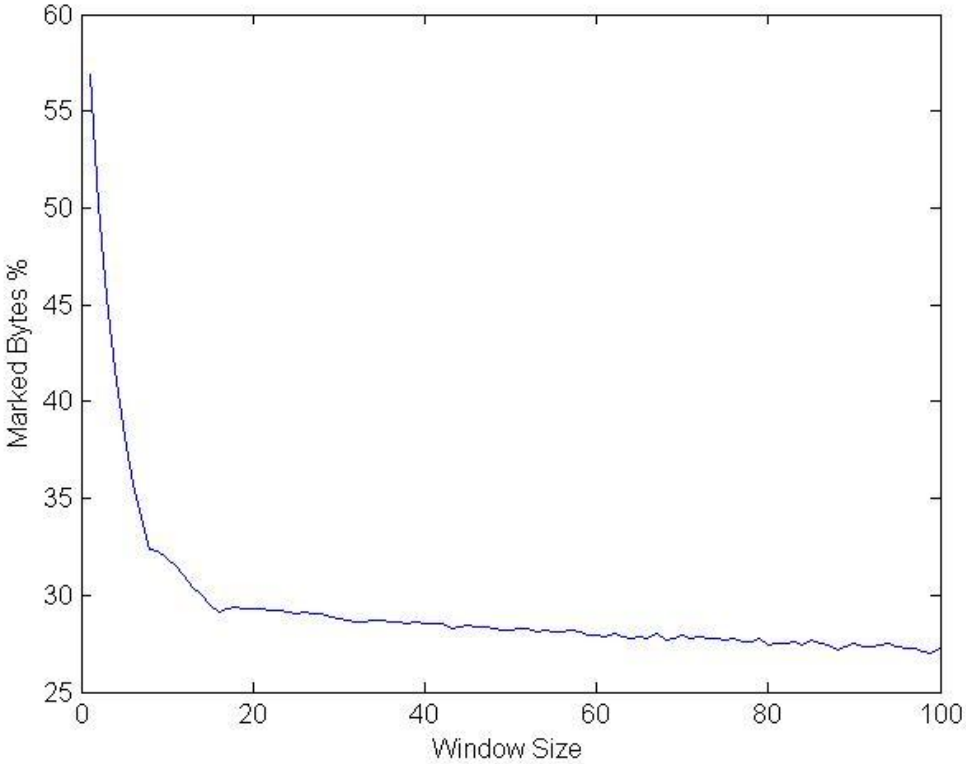

Figure 5. Silence of the Lambs B7, QP 38. Marked traffic versus the window size, with the use of the JW mechanism.

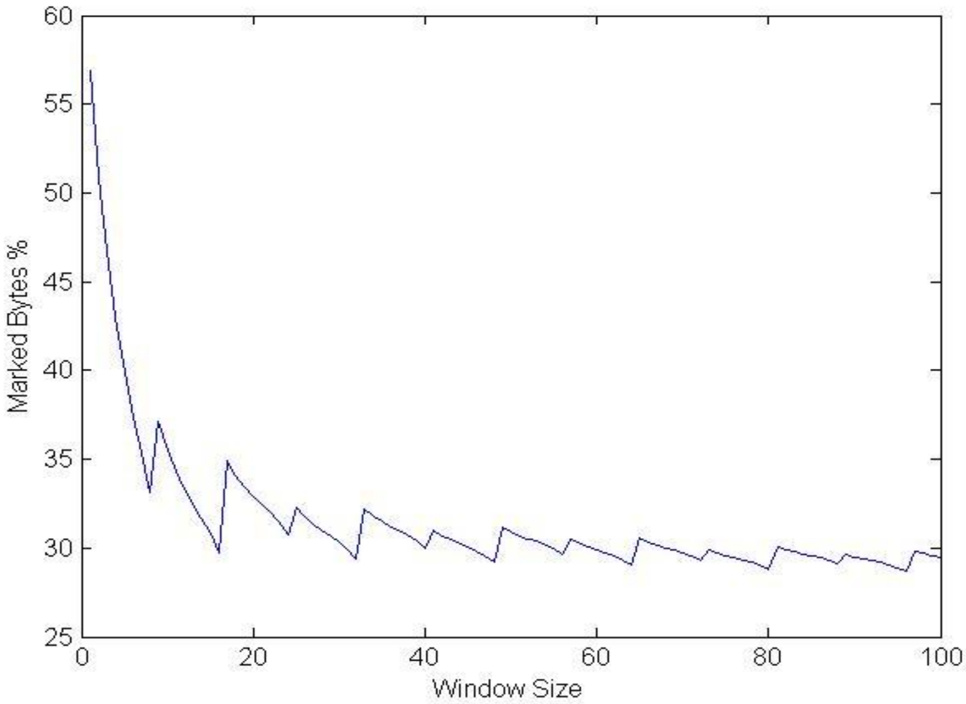

Figure 6. Silence of the Lambs, B7, QP 38. Marked traffic versus the window size, with the use of the MW mechanism.

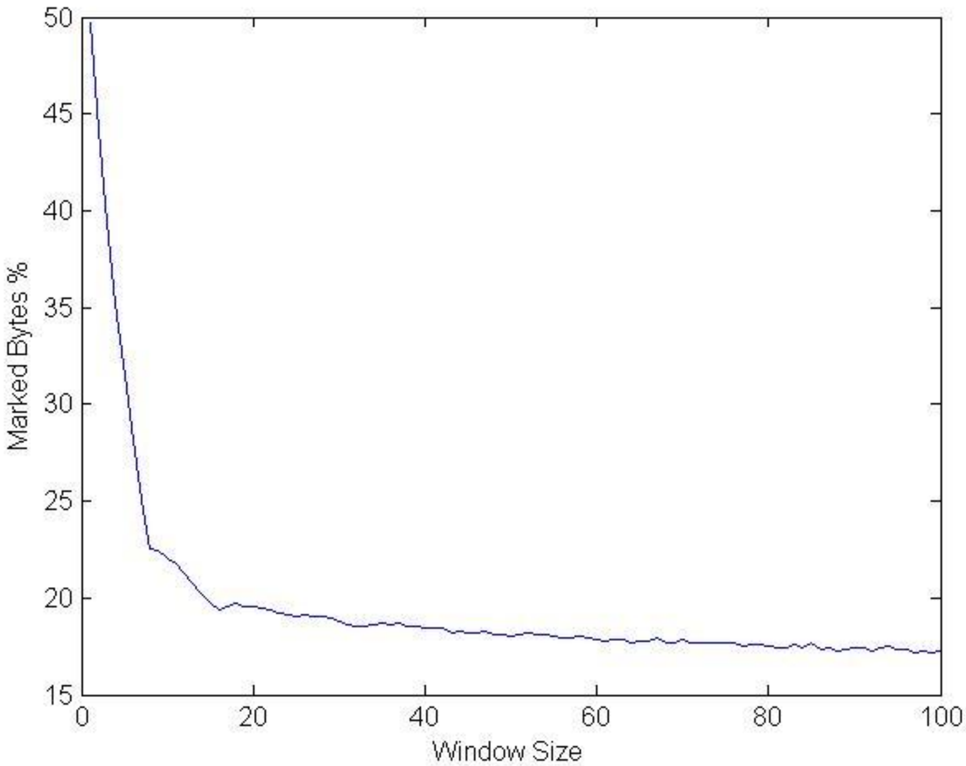

Figure 7. Silence of the Lambs, B7, QP48. Marked traffic versus the window size, with the use of the JW mechanism.

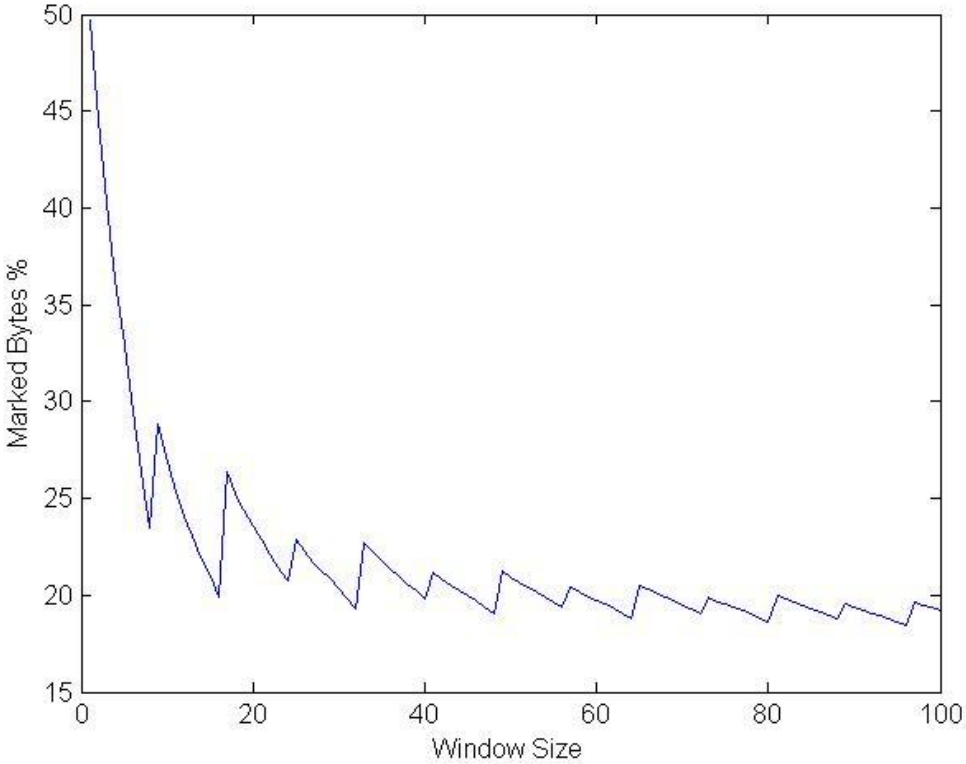

Figure 8. Silence of the Lambs, B7, QP 48. Marked traffic versus the window size, with the use of the MW mechanism.

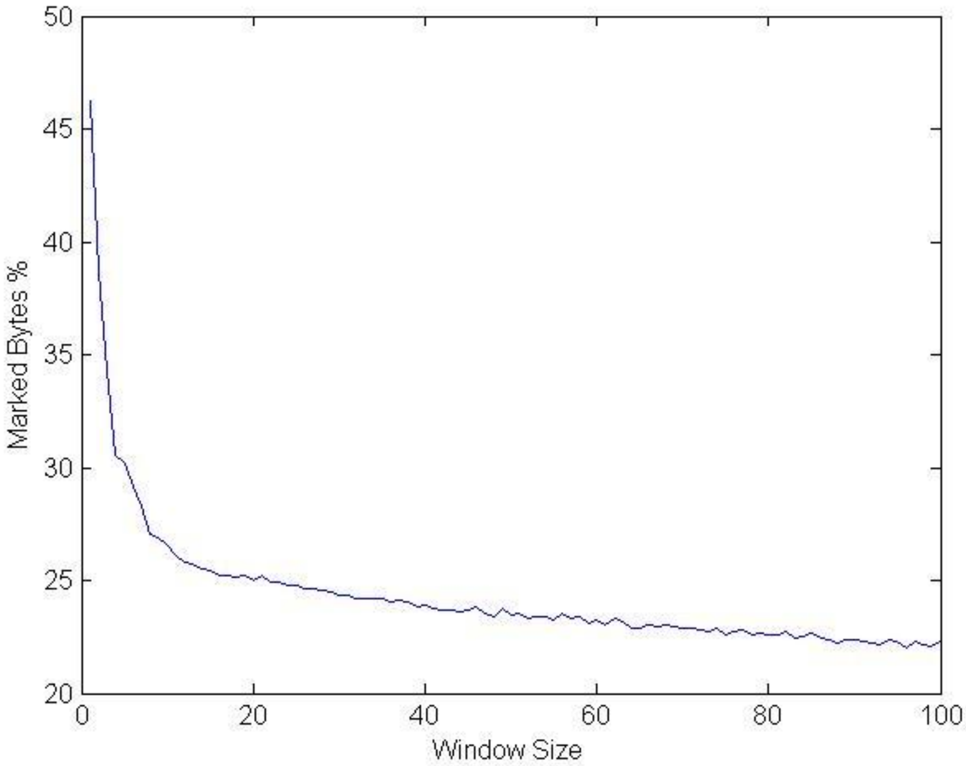

Figure 9. StarWars IV, B3, QP 16. Marked traffic versus the window size, with the use of the JW mechanism.

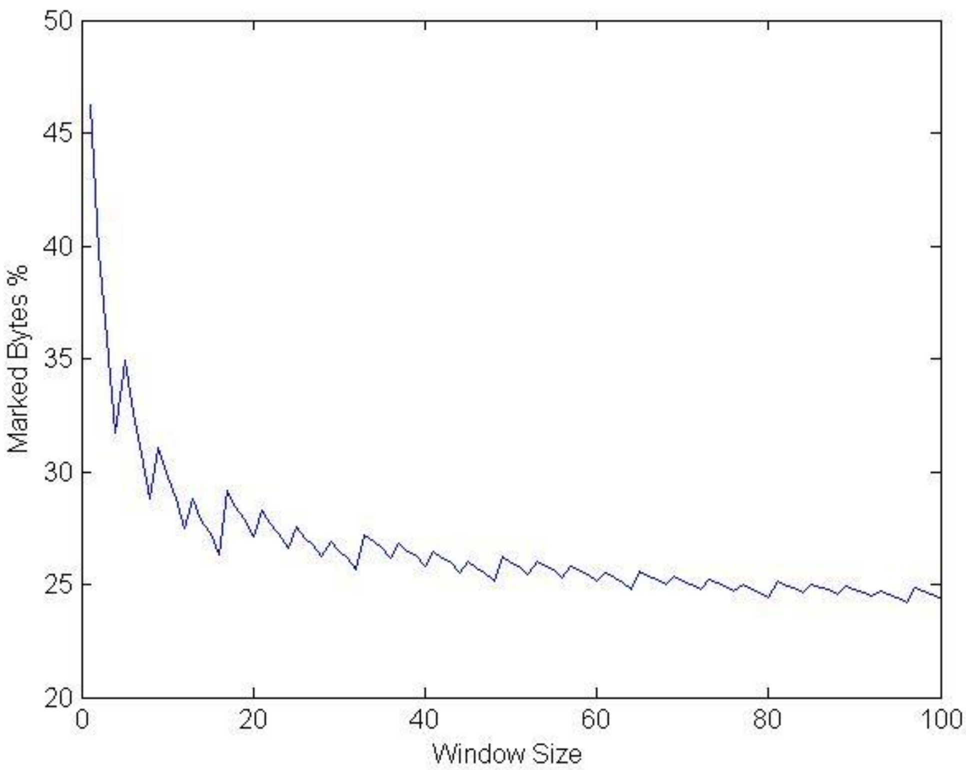

Figure 10. StarWars IV, B3, QP 16. Marked traffic versus the window size, with the use of the MW mechanism.

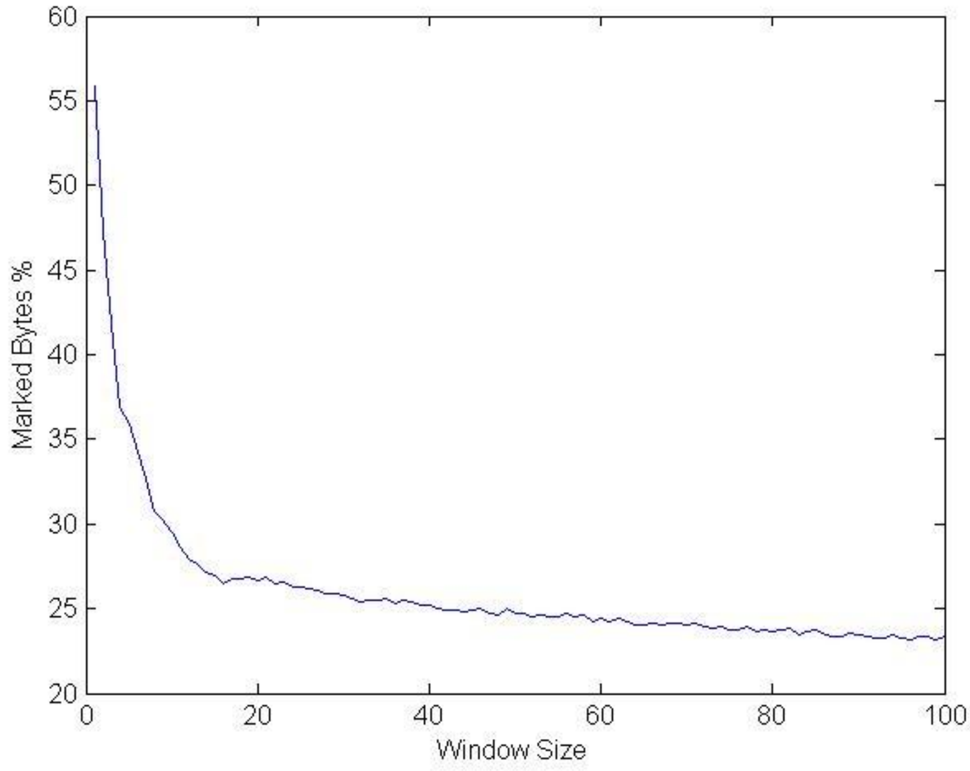

Figure 11. StarWars IV, B3, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

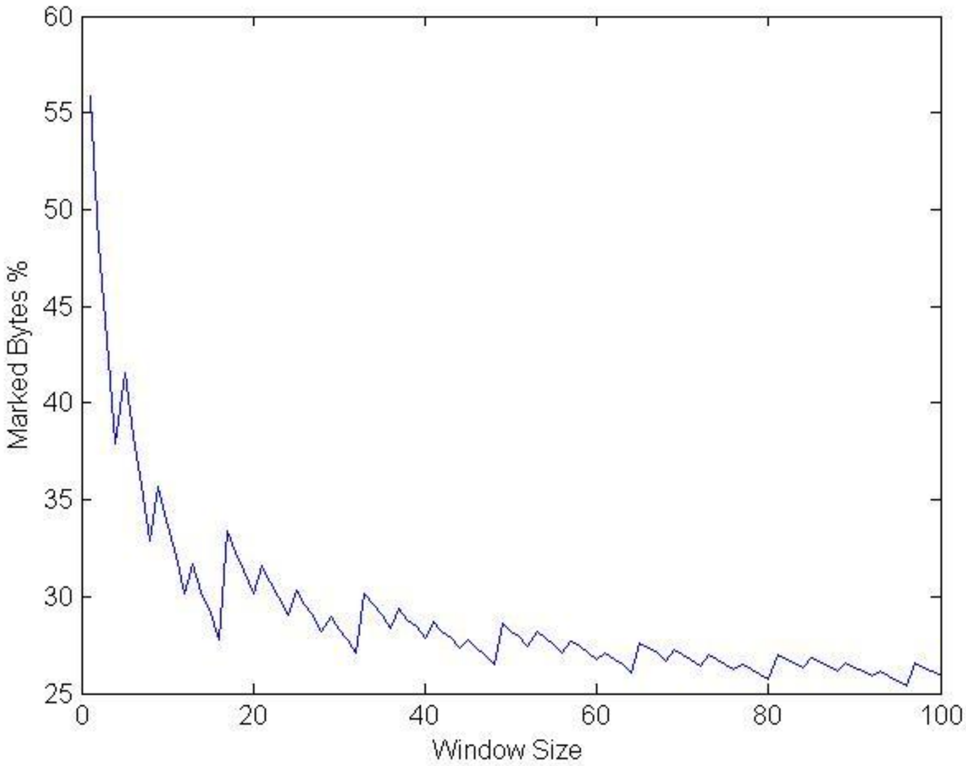

Figure12. StarWars IV, B3, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

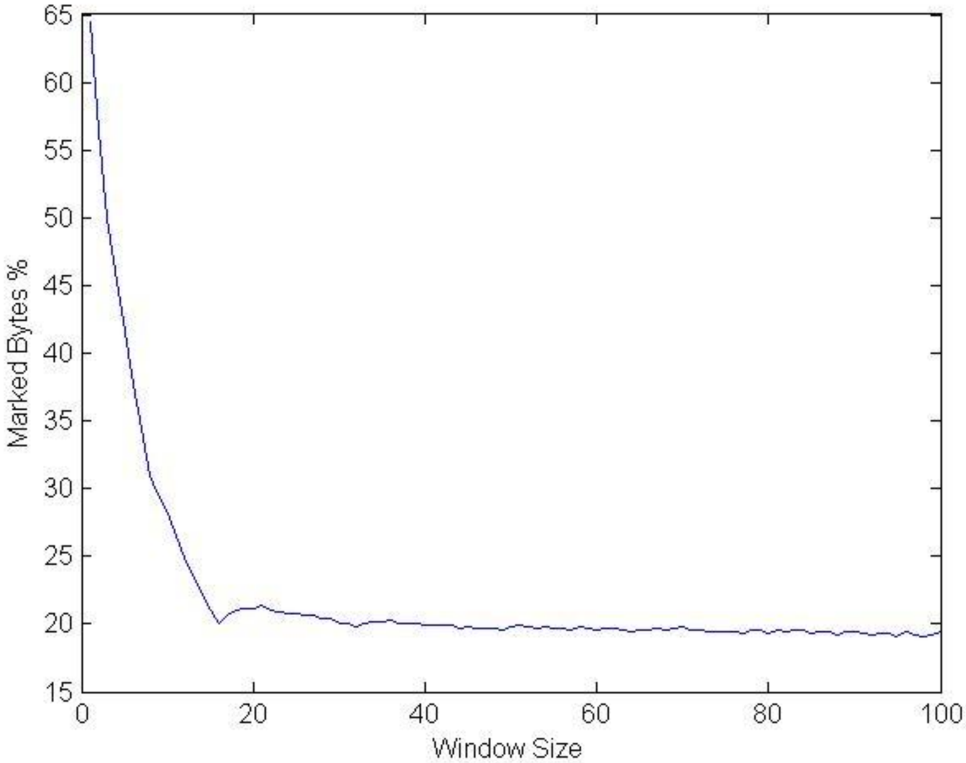

Figure13. SonyDemo, B7, QP 38. Marked traffic versus the window size, with the use of the JW mechanism.

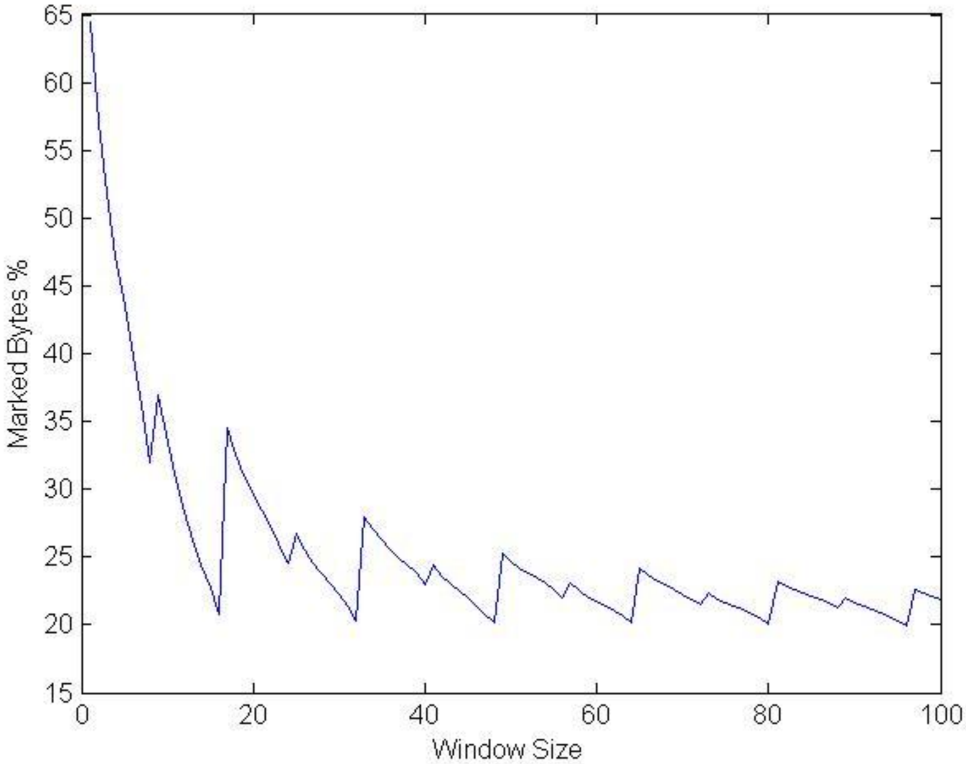

Figure14. SonyDemo, B7, QP 38. Marked traffic versus the window size, with the use of the MW mechanism.

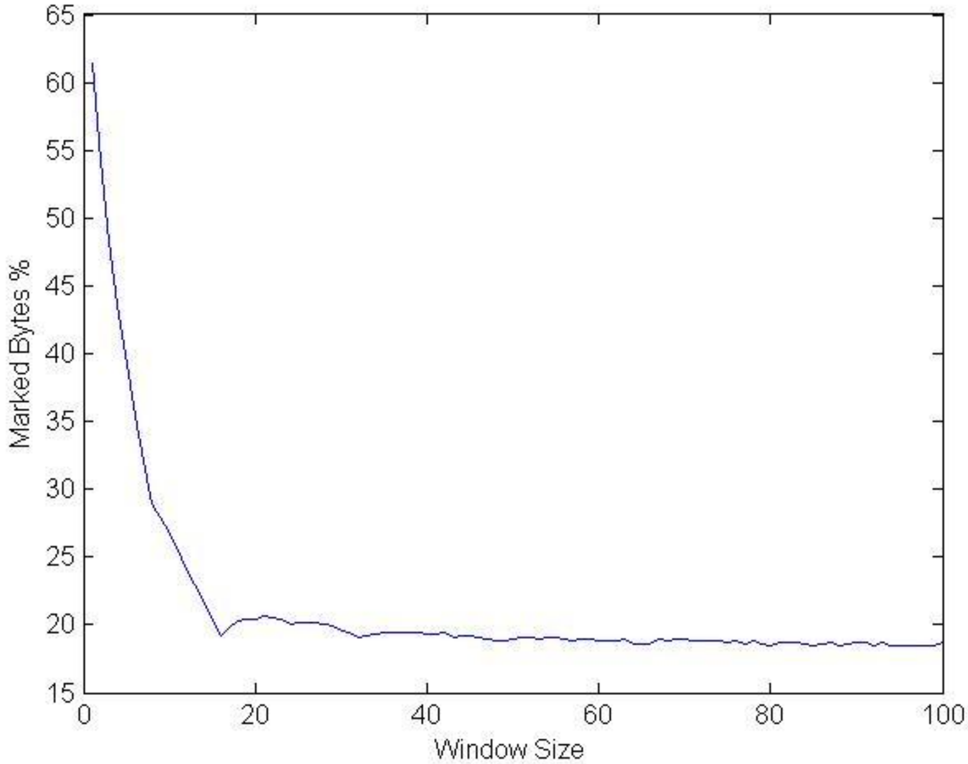

Figure15. SonyDemo, B7, QP 48. Marked traffic versus the window size, with the use of the JW mechanism.

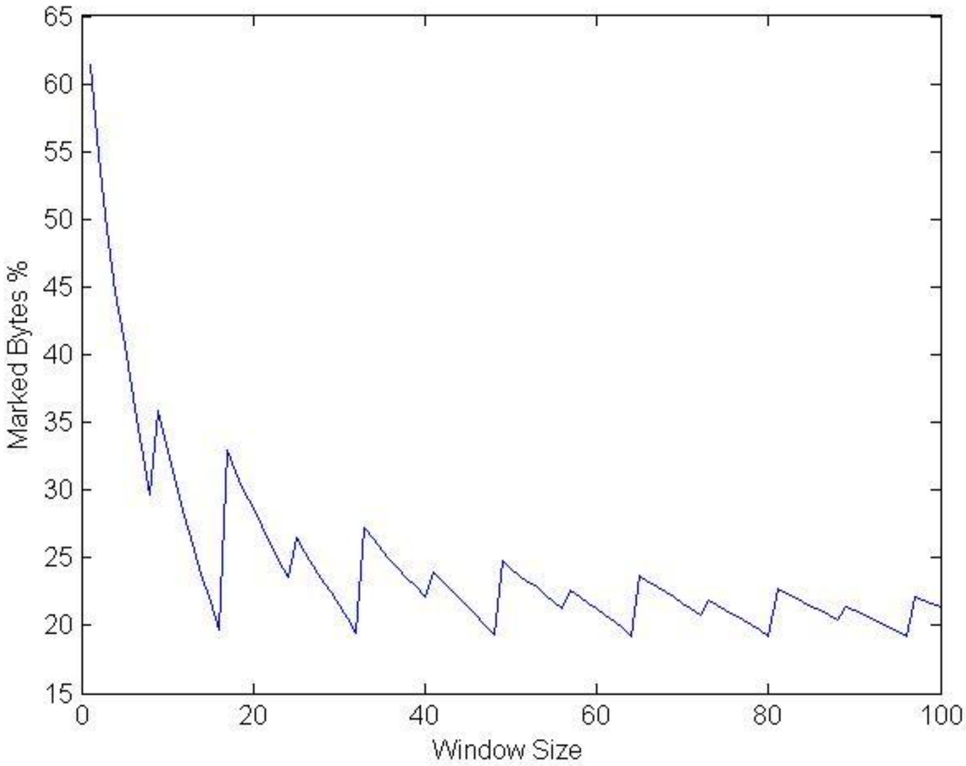

Figure16. SonyDemo, B7, QP 48. Marked traffic versus the window size, with the use of the MW mechanism.

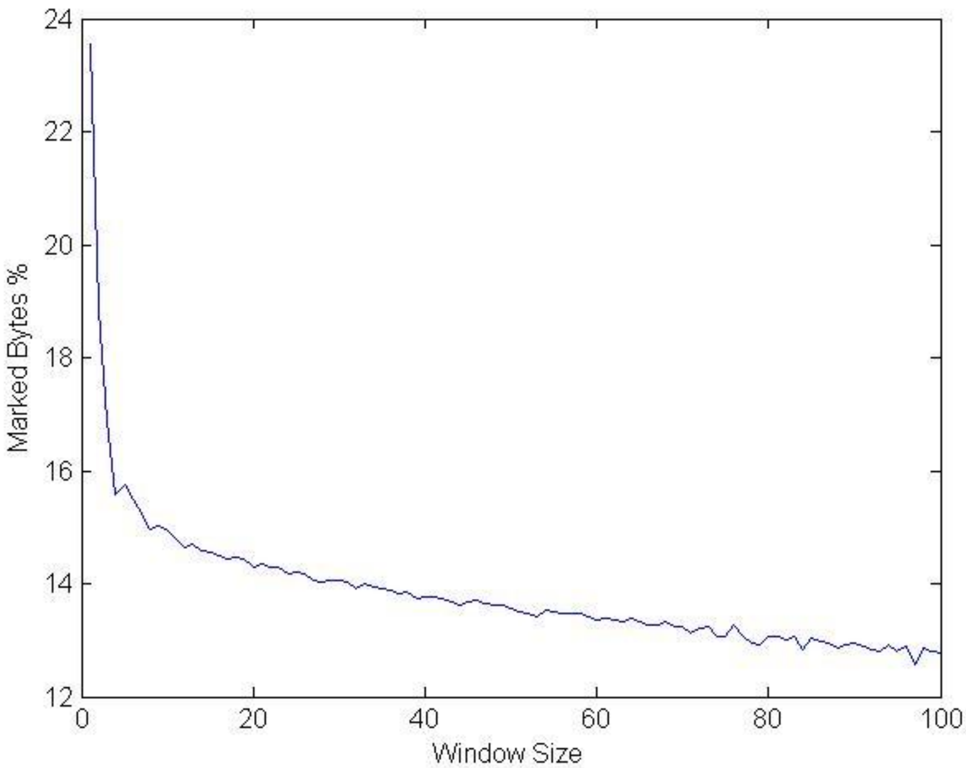

Figure17. NbcNews, B3, QP 16. Marked traffic versus the window size, with the use of the JW mechanism.

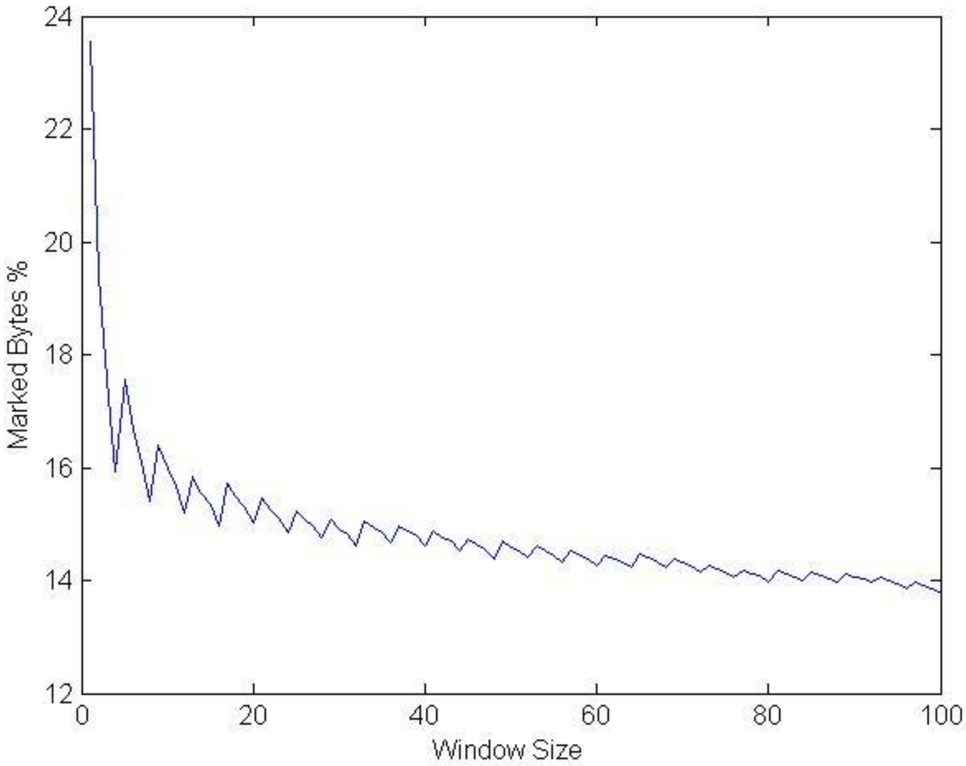

Figure18. NbcNews, B3, QP 16. Marked traffic versus the window size, with the use of the MW mechanism.

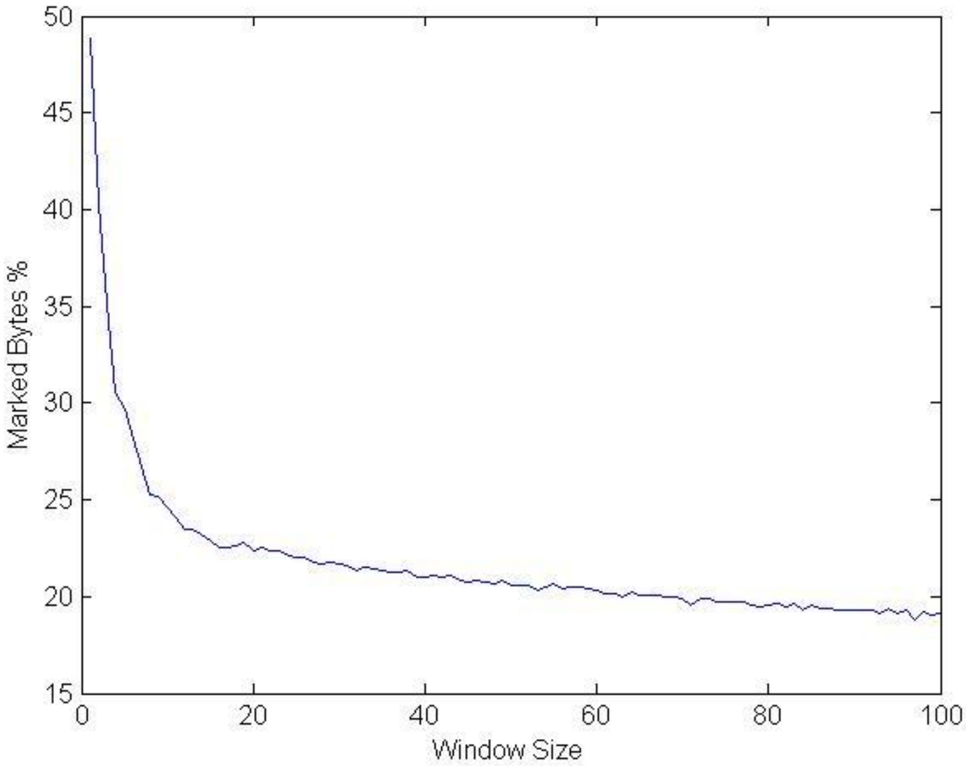

Figure19. NbcNews, B3, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

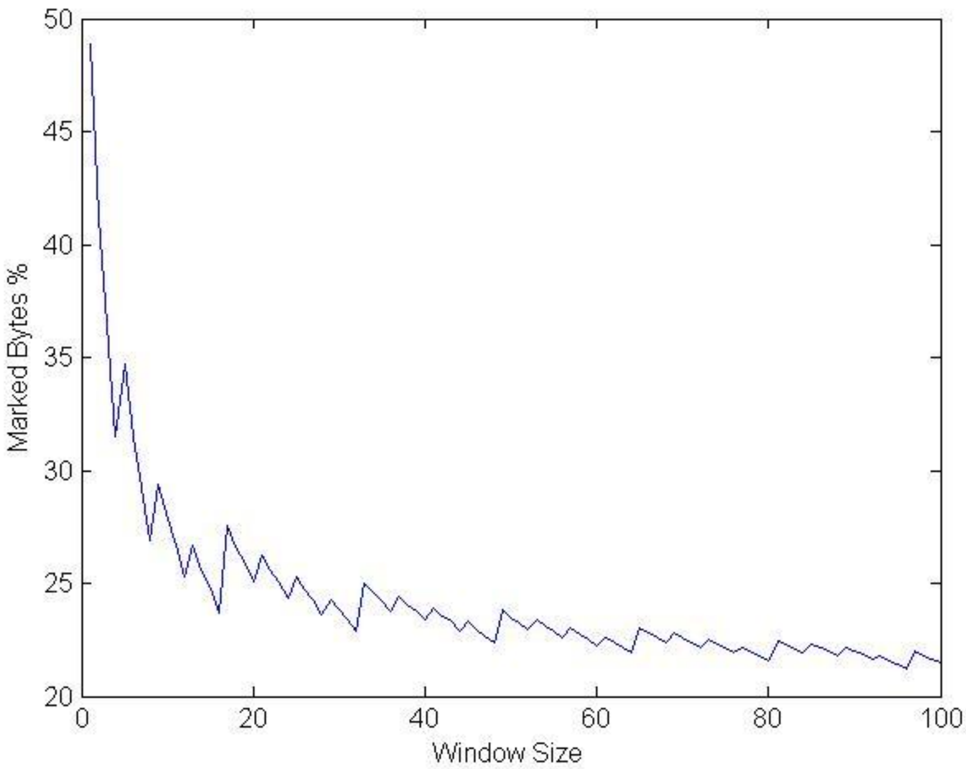

Figure20. NbcNews, B3, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

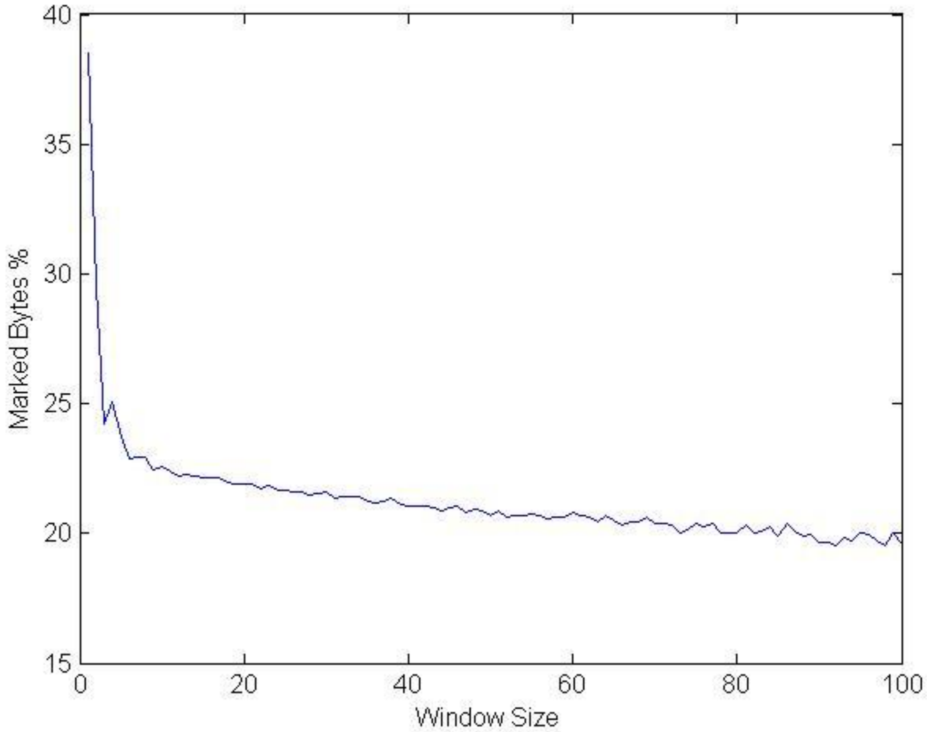

Figure21. Terminator, B2, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

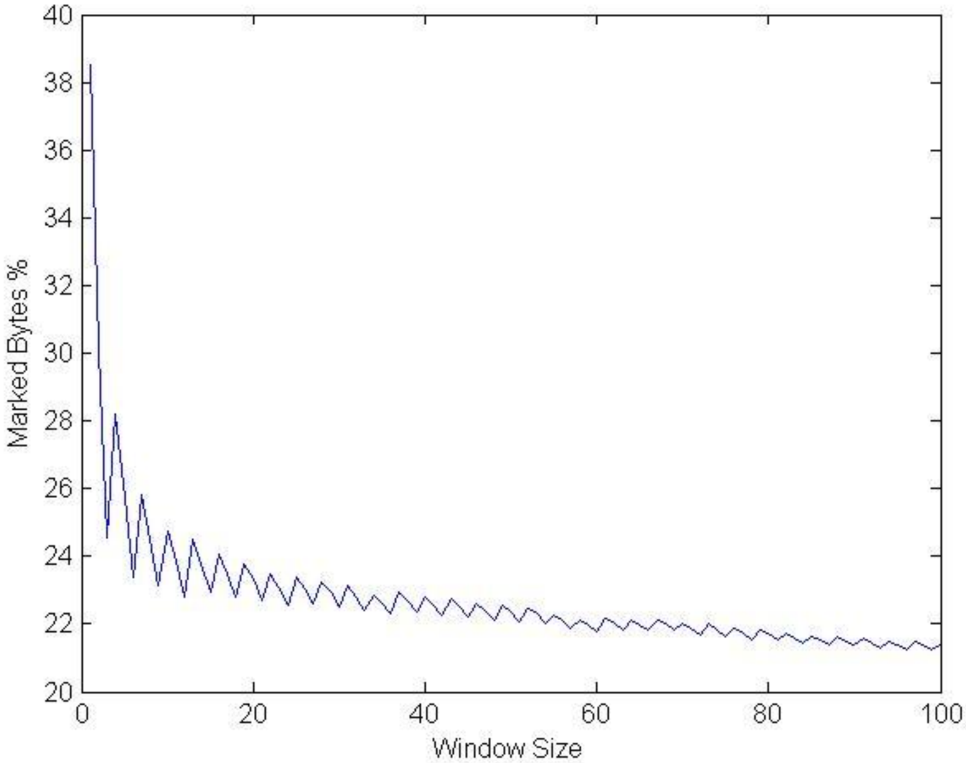

Figure22. Terminator, B2, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

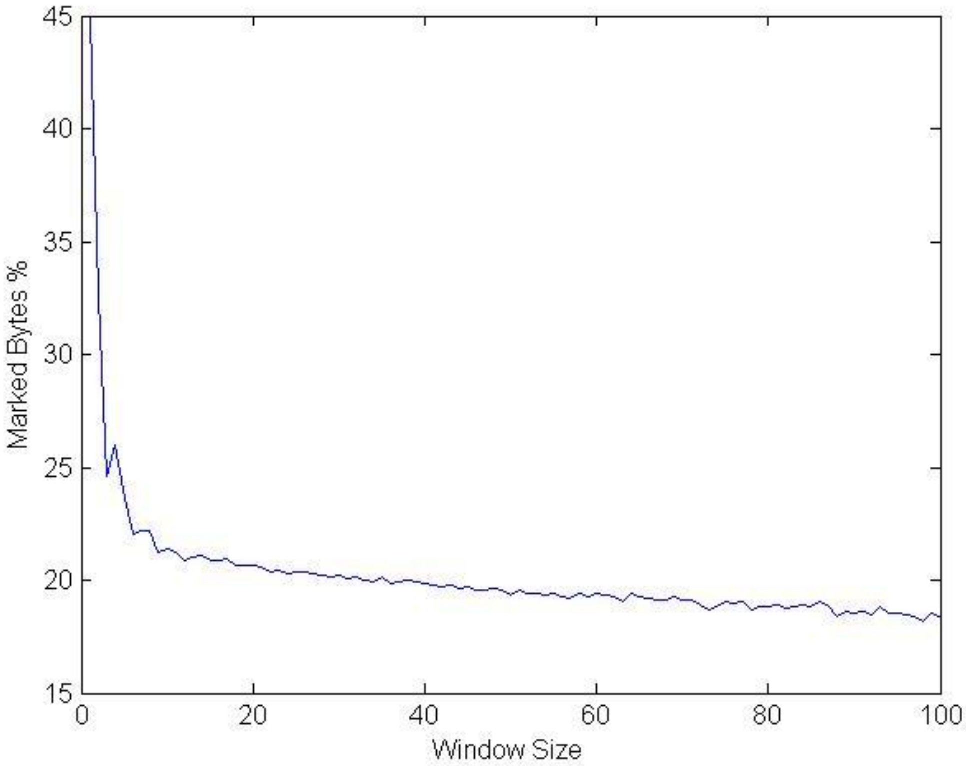

Figure23. Terminator, B2, QP 38. Marked traffic versus the window size, with the use of the JW mechanism.

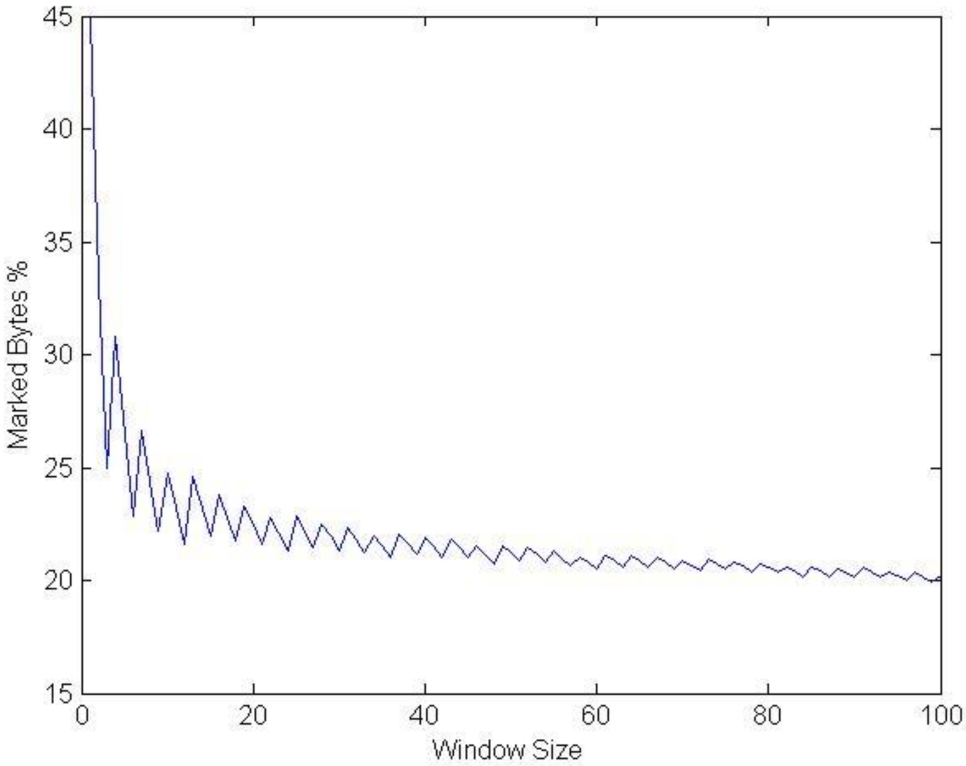

Figure24. Terminator, B2, QP 38. Marked traffic versus the window size, with the use of the MW mechanism.

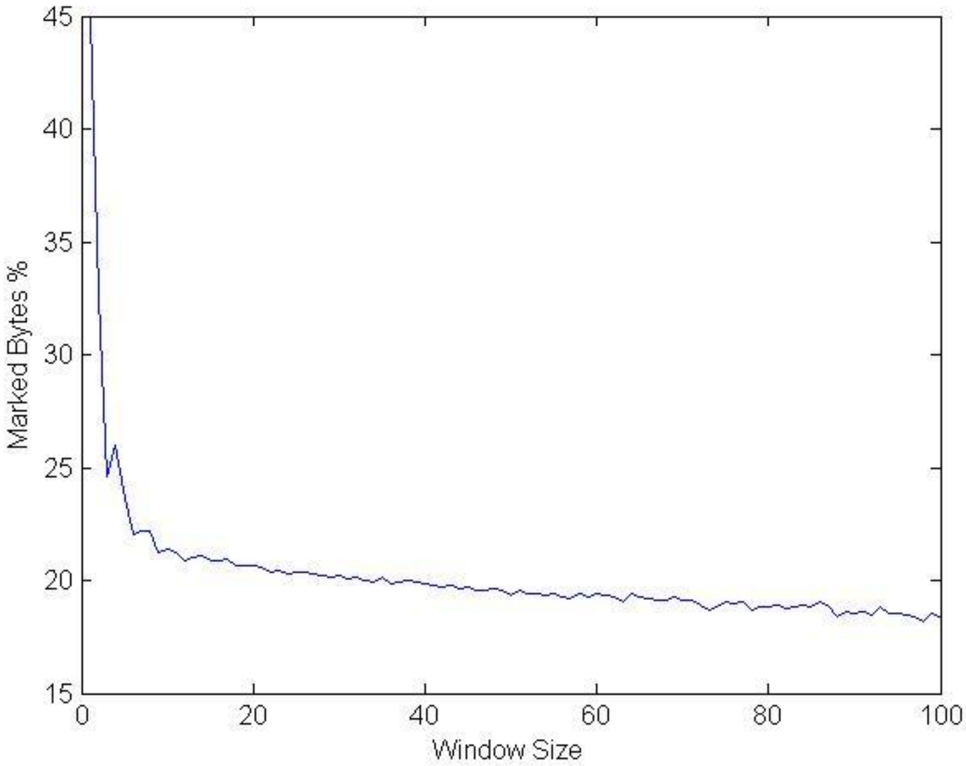

Figure25. Terminator, B2, QP 48. Marked traffic versus the window size, with the use of the JW mechanism.

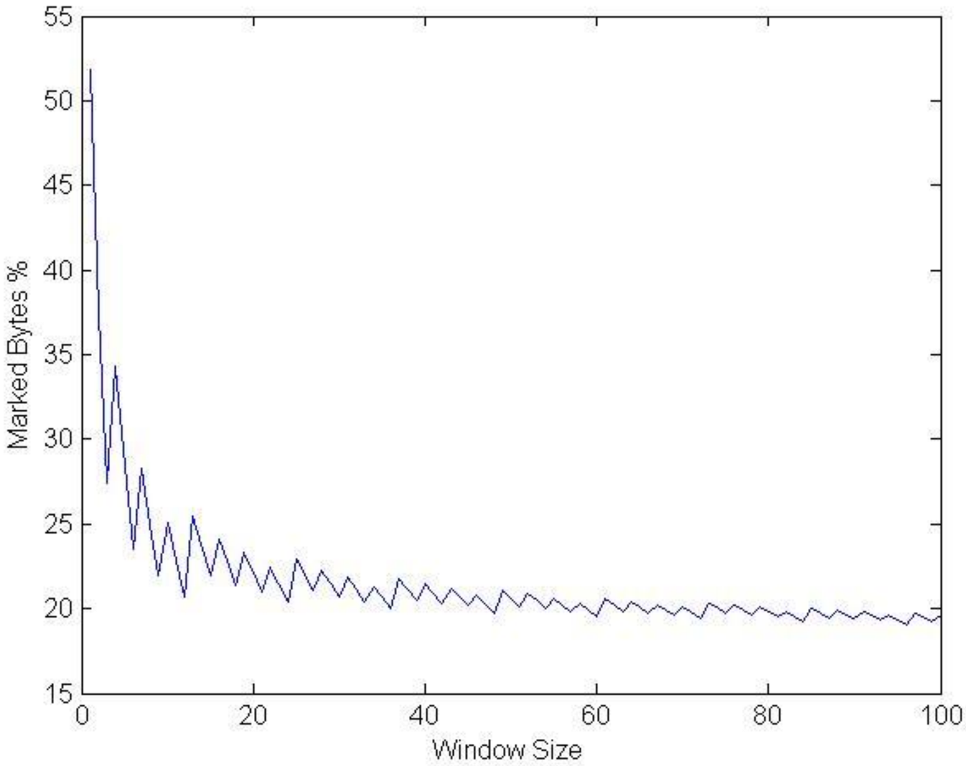

Figure26. Terminator, B2, QP 48. Marked traffic versus the window size, with the use of the MW mechanism.

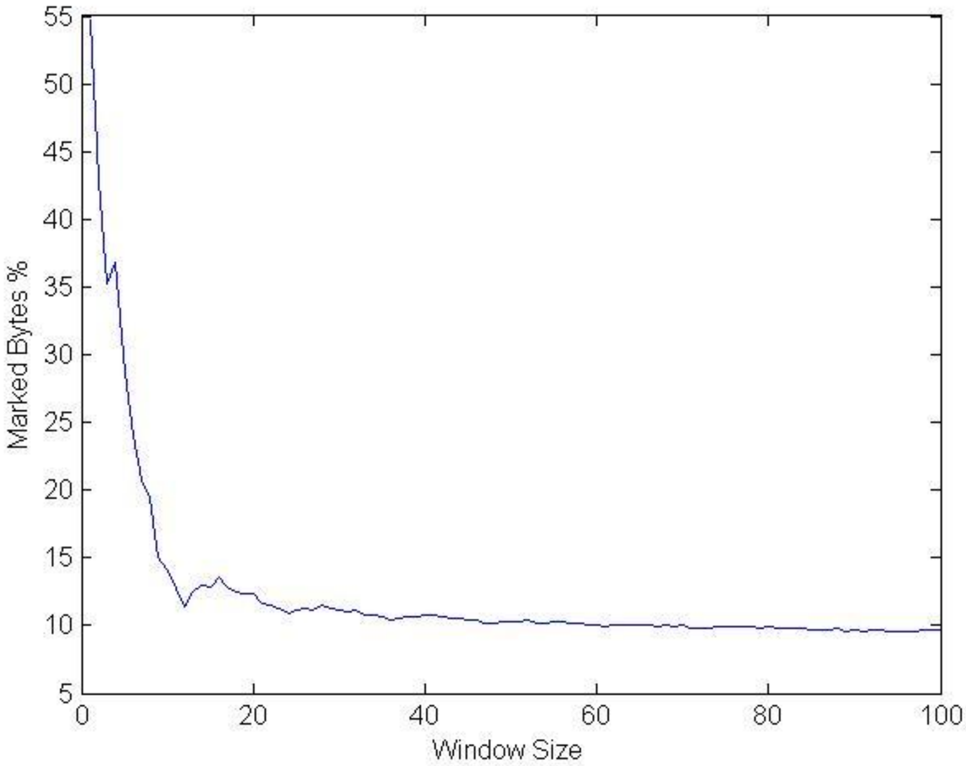

Figure27. Kaet's Horizon, B2, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

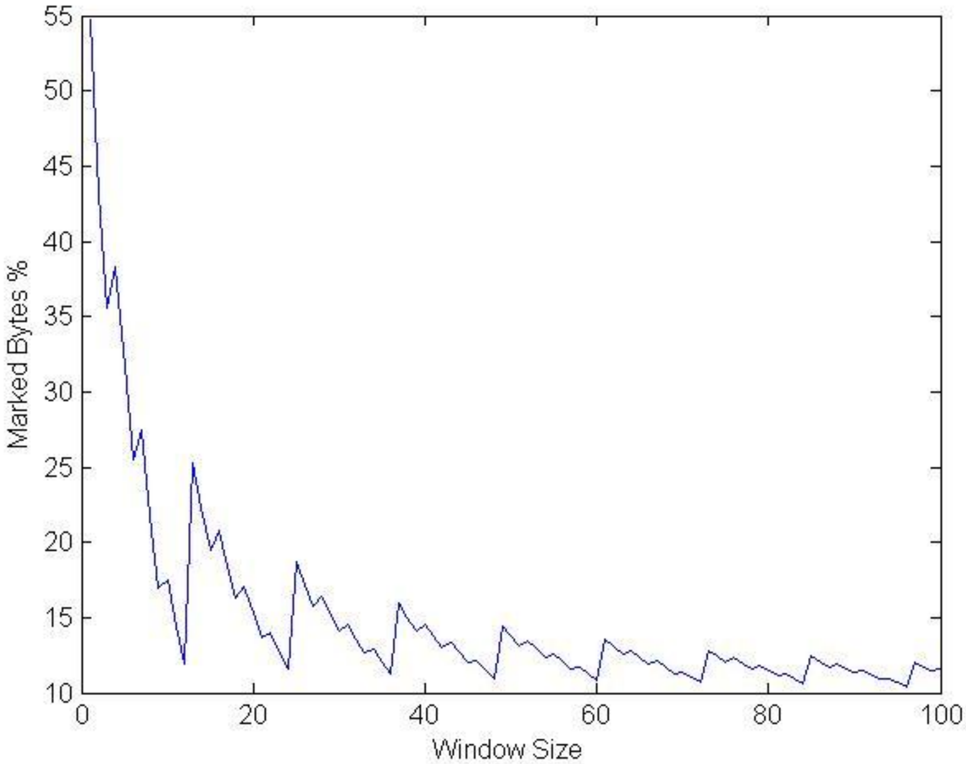

Figure28. Kaet's Horizon, B2, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

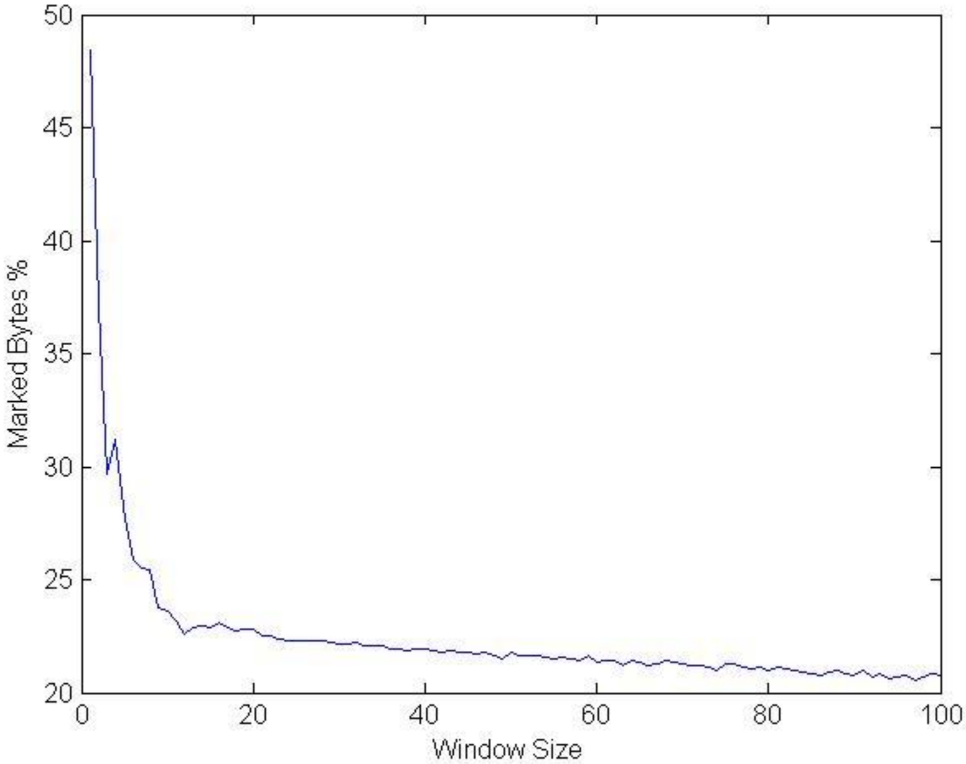

Figure29. Kaet's from Mars to China, B2, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

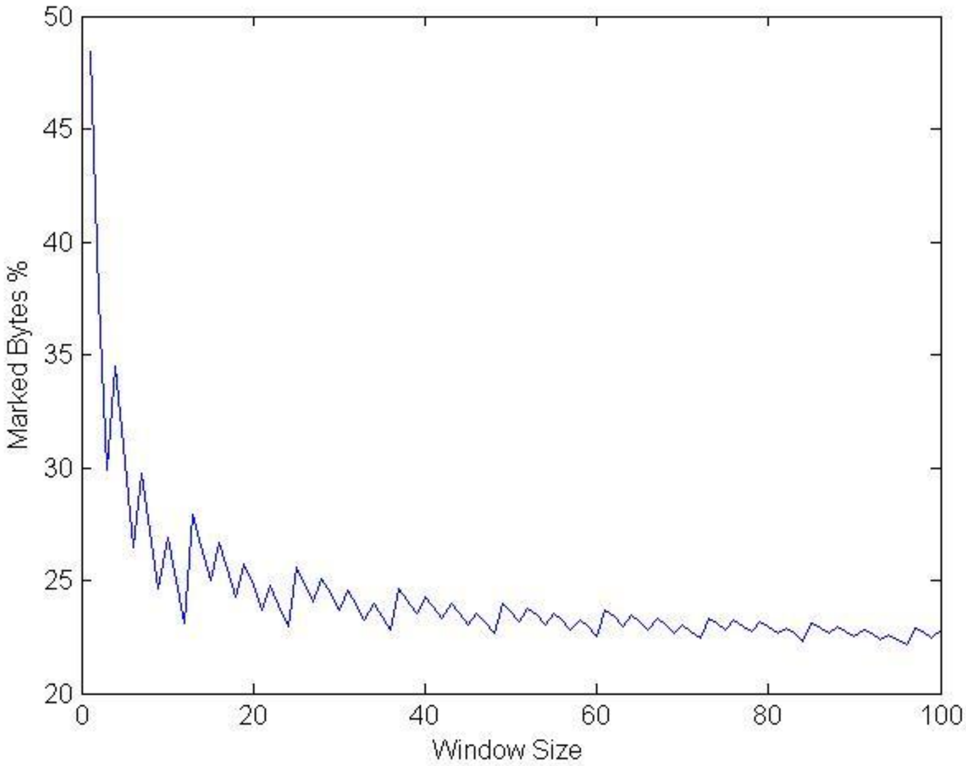

Figure30. Kaet's from Mars to China, B2, QP 28. Marked traffic versus the window size, with the use of the MW mechanism.

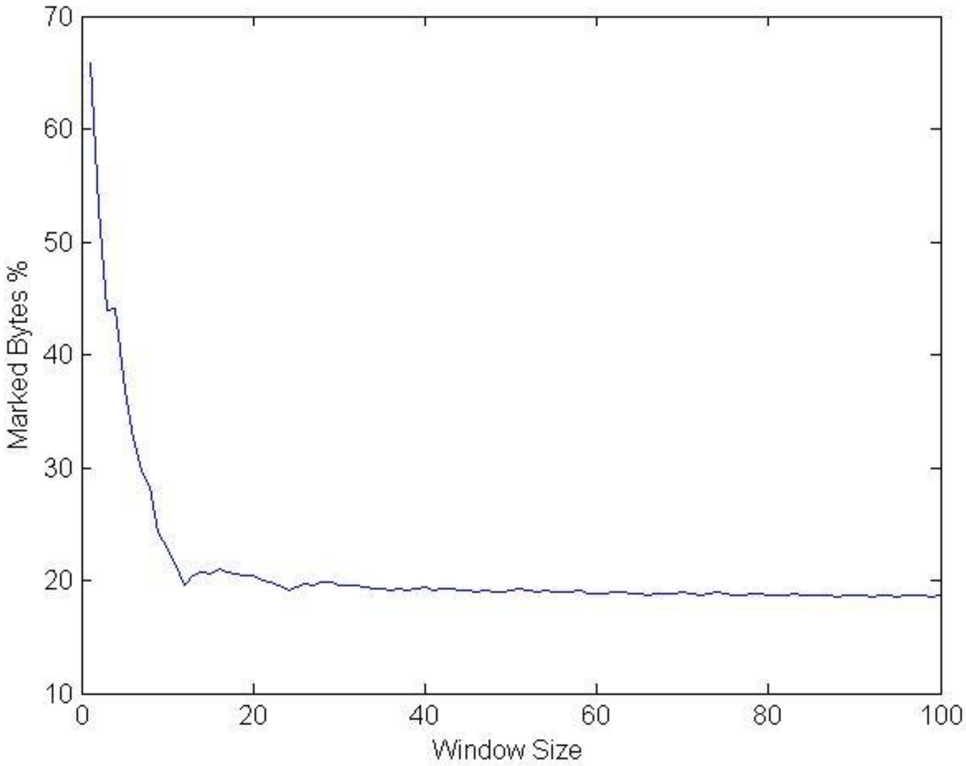

Figure31. Sonydemo2, B2, QP 28. Marked traffic versus the window size, with the use of the JW mechanism.

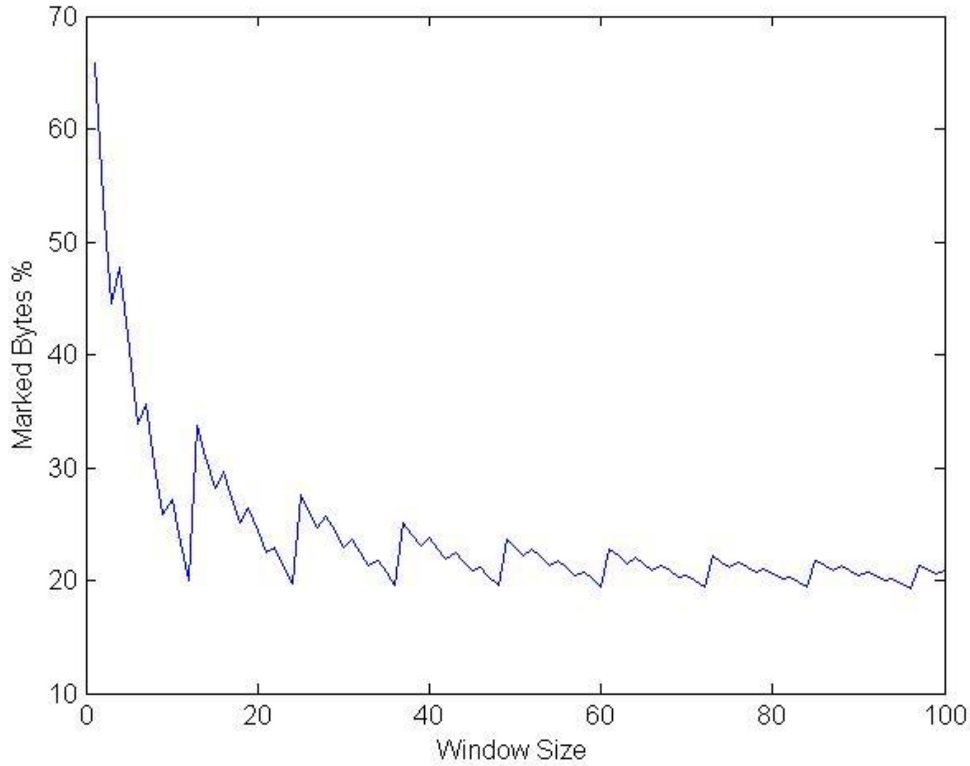

Figure 32. Sonydemo2, B2, QP 28 Marked traffic versus the window size, with the use of the MW mechanism.

4.4. Results with the use of EWMA

We derived our results with the use of 5 different g values: 0.1, 0.3, 0.5, 0.7 and 0.9, for 4 different window sizes, equal to the duration of 5, 30, 55 and 80 video frames, respectively. As shown in Figures 33-46, the percentage of marked traffic decreases as the window length and the value of g increase. In order to find out which of the two parameters (window size, g value) is the most influential in the decrease of marked traffic, we performed two additional sets of simulations. In the first, we kept the window size fixed and we changed the value of g, progressively, from 0.1 to 0.9 in increments of 0.2. We conducted simulations both for a small fixed window size (duration equal to 5 video frames) and a large fixed window size (duration equal to 80 video frames). The average decrease in the percentage of the marked traffic, for all values of g, was 2.23% for the small window size and 1.15% for the large window size. On the contrary, when we kept the value of g fixed (to 0.1, 0.3, 0.5, 0.7 and 0.9, respectively) and changed the window size, from N=5 to N=80, in increments of 25, the average decrease in the percentage of marked traffic over all the simulated scenarios was 7.49%. Therefore, it is clear that although both the value of g and the window size can help decrease the marked traffic of conforming video users, the window size is the more influential of the two.

Still, even for g=0.9 and a window of size equal to the duration of 80 video frames (i.e., even with a choice of very lenient traffic policing parameters) the percentage of marked traffic remains very high, therefore the probability that the conforming user will enjoy good QoS is slim (the marked traffic will be dropped in the first instance of network congestion). Hence, we conclude once again that the burstiness of H.264 traffic calls for more lenient policing, in terms of the allowed packet transmission rate per video frame. This rate needs to be larger than the declared mean rate of the source, in order to avoid the violation of the QoS requirements of conforming video users.

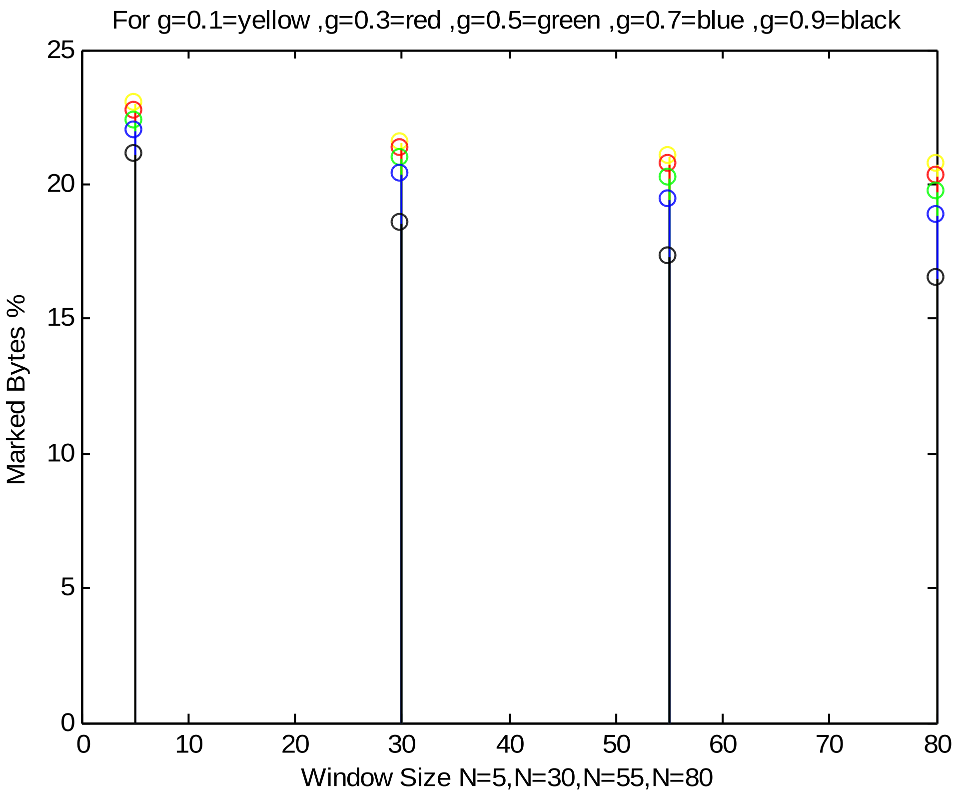

Figure 33. Tokyo Olympics B3, QP16. Marked traffic with the use of the EWMA mechanism.

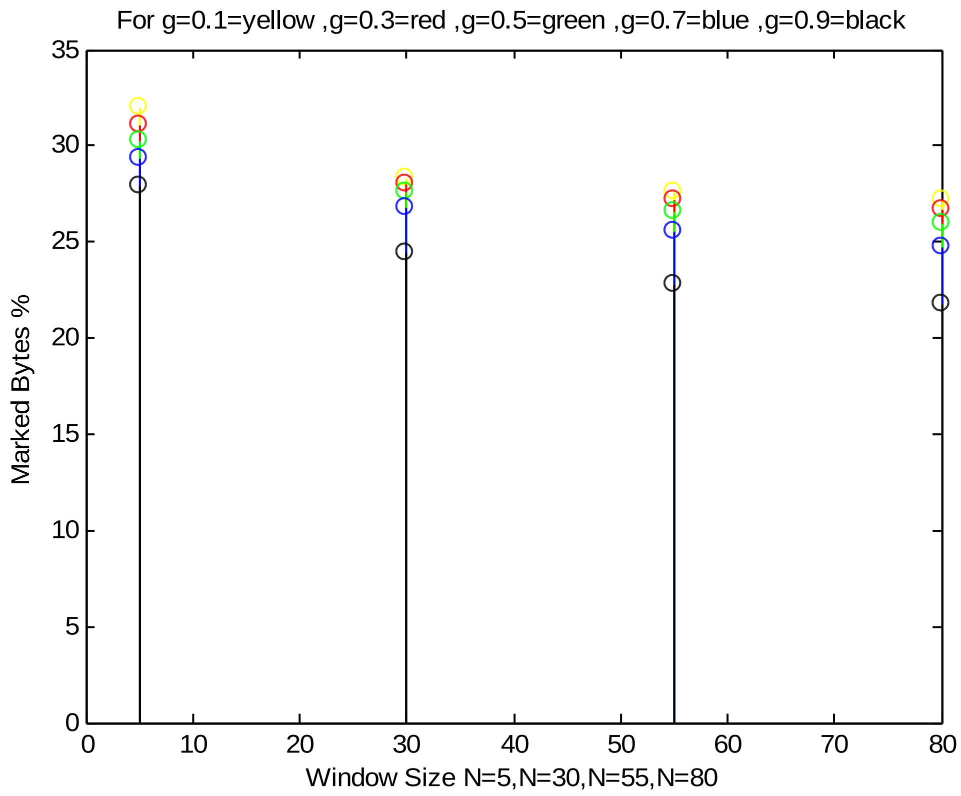

Figure 34. Tokyo Olympics B3, QP28. Marked traffic with the use of the EWMA mechanism.

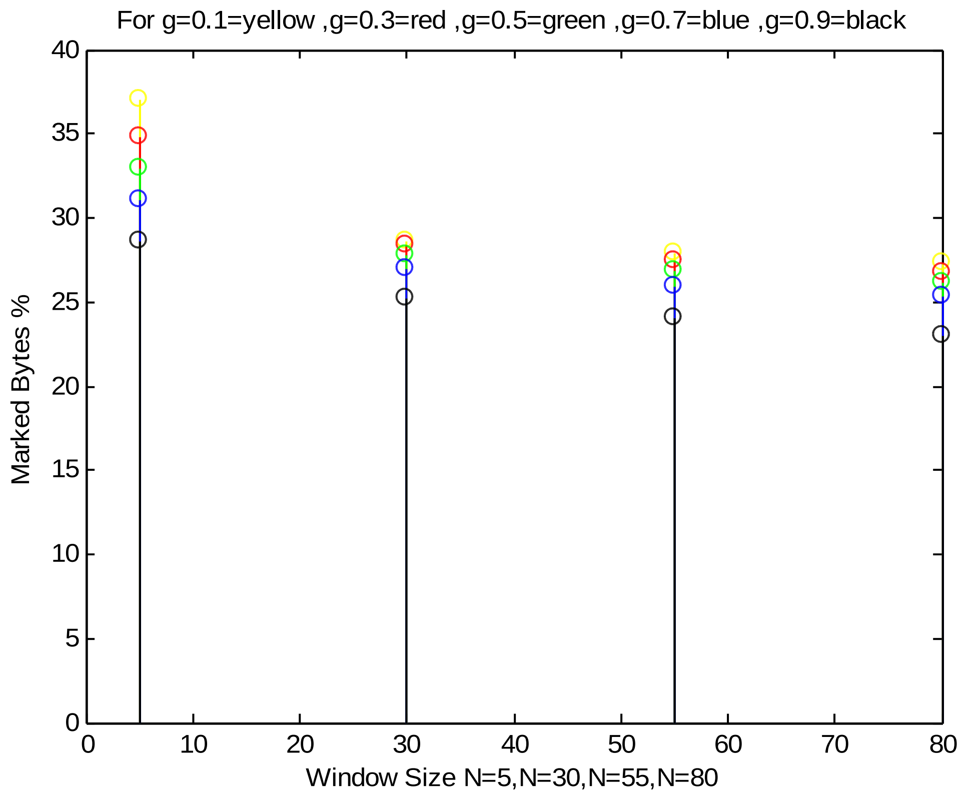

Figure 35. Silence of the Lambs, B7, QP38. Marked traffic with the use of the EWMA mechanism.

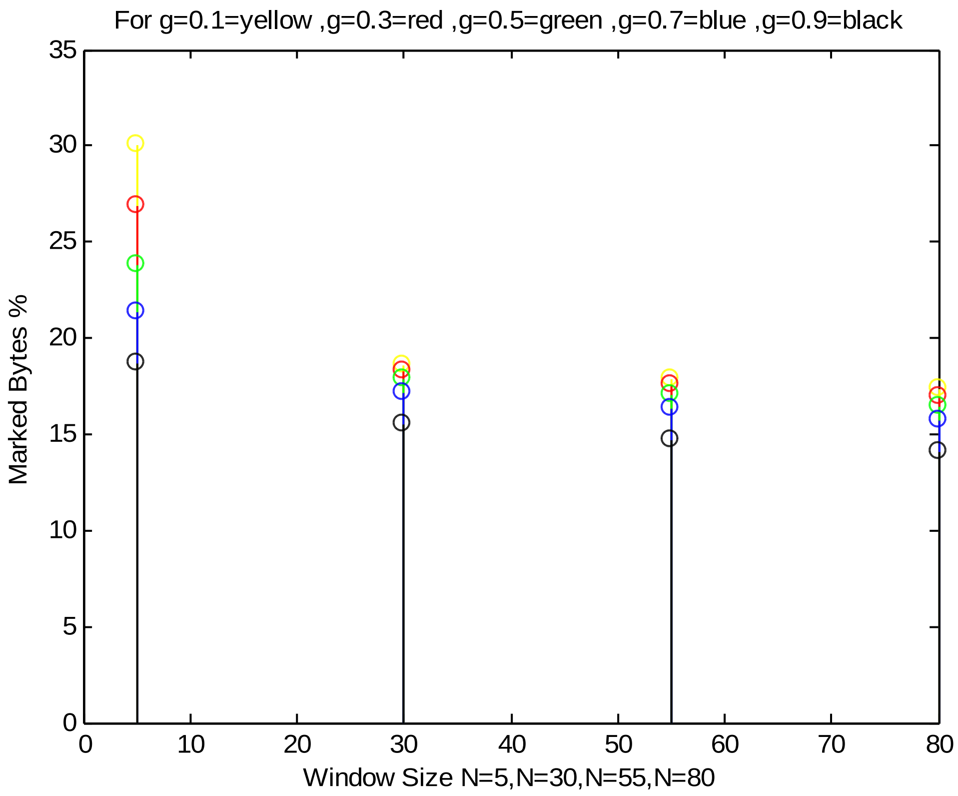

Figure 36. Silence of the Lambs, B7, QP48. Marked traffic with the use of the EWMA mechanism.

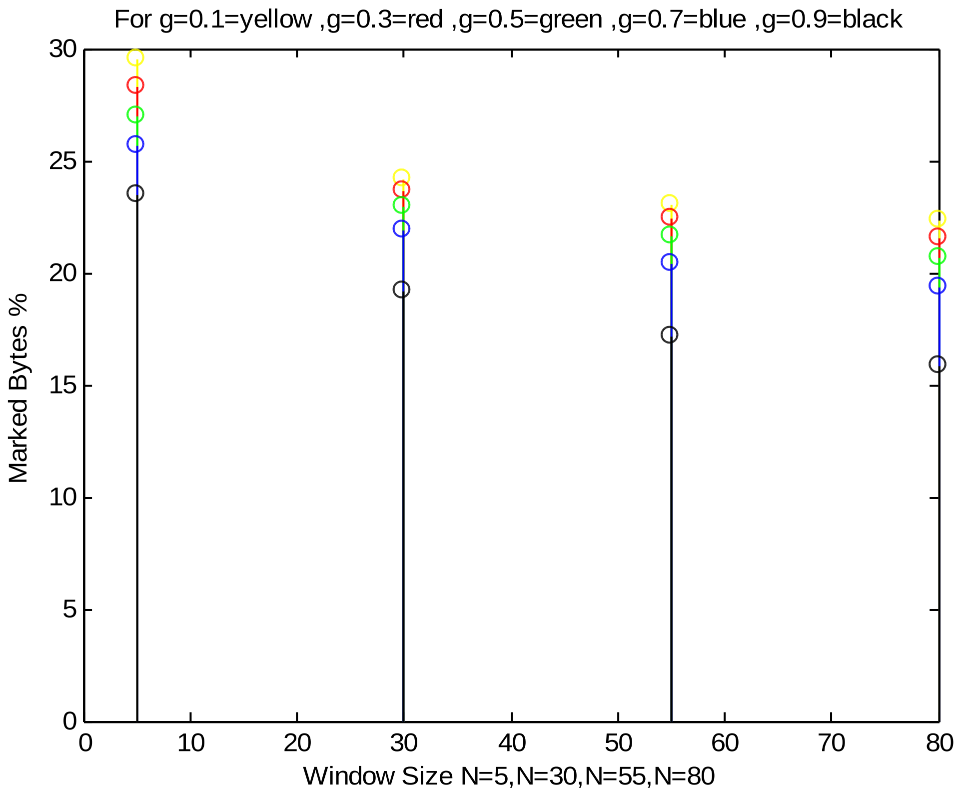

Figure 37. StarWars IV, B3, QP16. Marked traffic with the use of the EWMA mechanism.

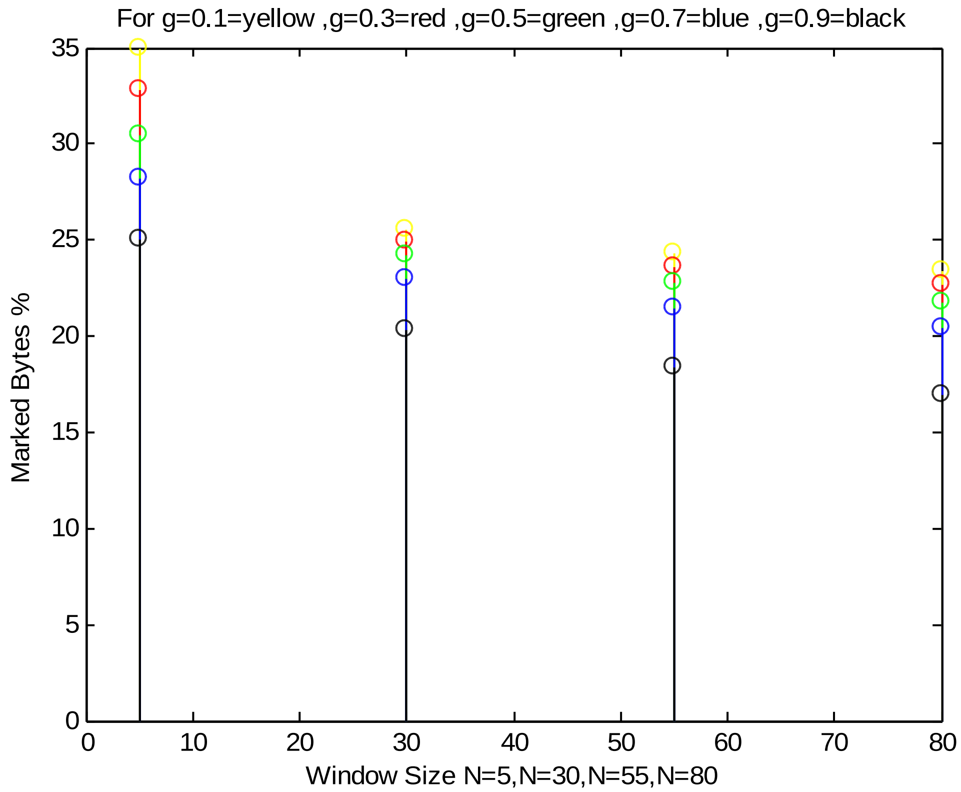

Figure 38. StarWars IV, B3, QP28. Marked traffic with the use of the EWMA mechanism.

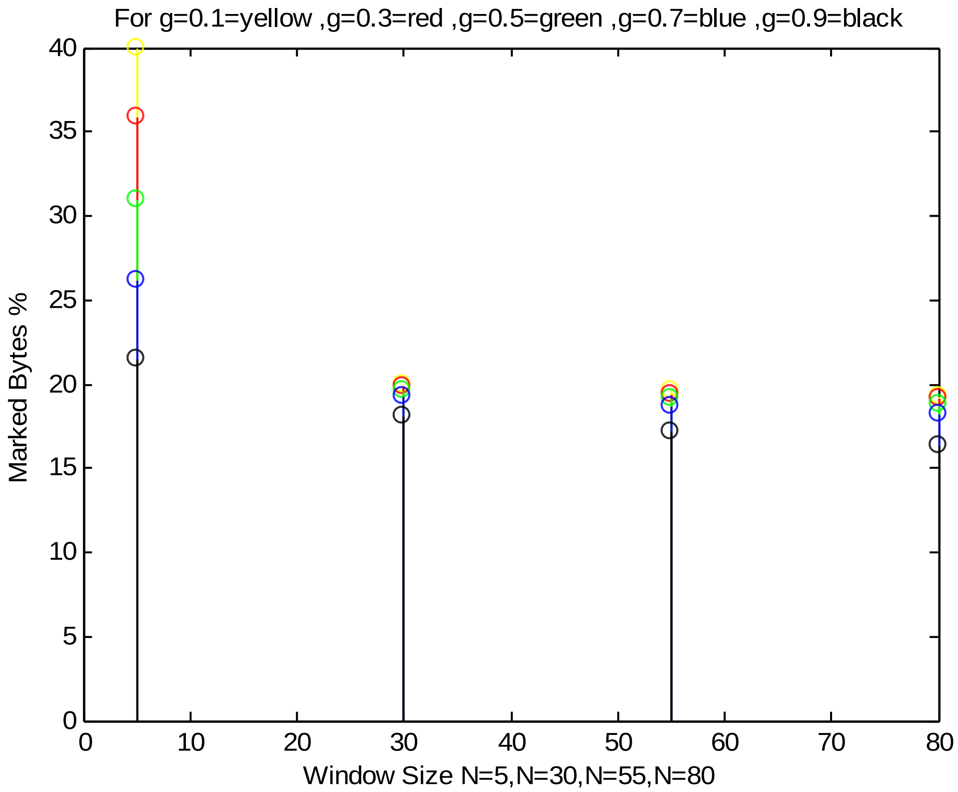

Figure 39. SonyDemo, B7, QP38. Marked traffic with the use of the EWMA mechanism.

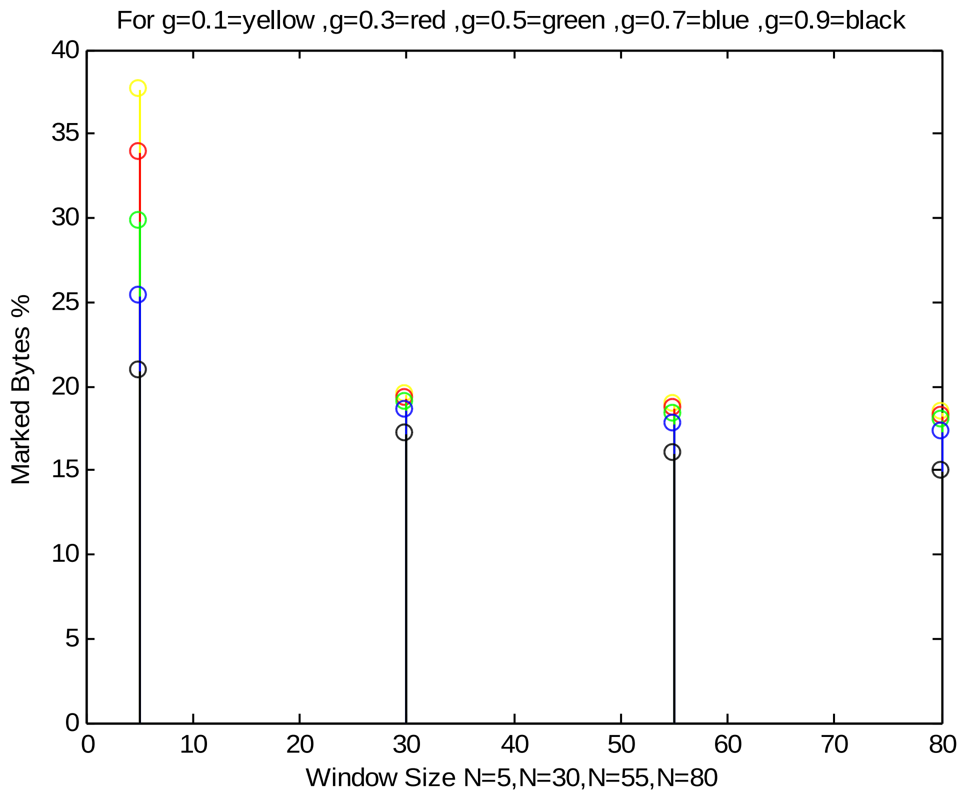

Figure 40. SonyDemo, B7, QP48. Marked traffic with the use of the EWMA mechanism.

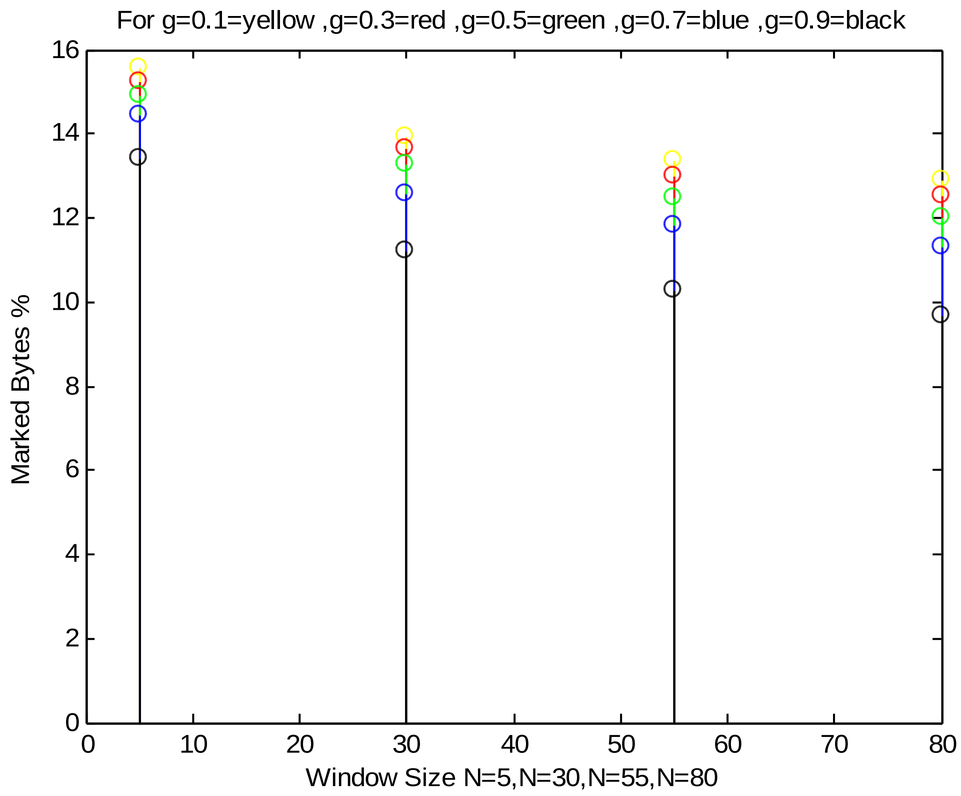

Figure 41. NbcNews, B3, QP16. Marked traffic with the use of the EWMA mechanism.

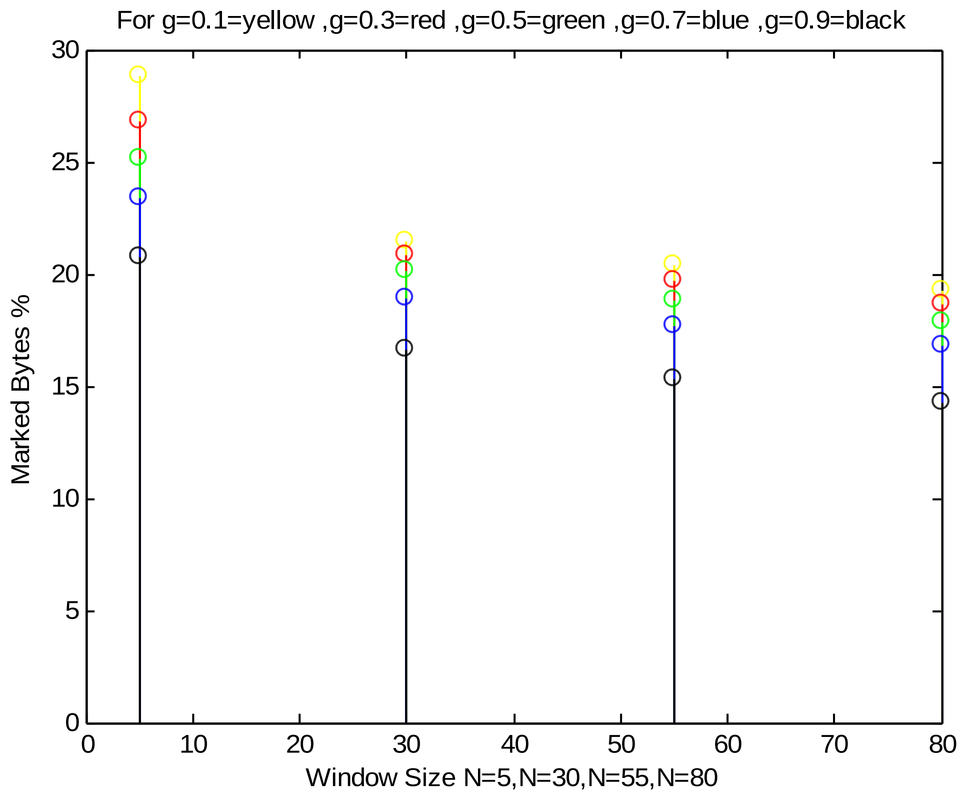

Figure 42. NbcNews, B3, QP28. Marked traffic with the use of the EWMA mechanism.

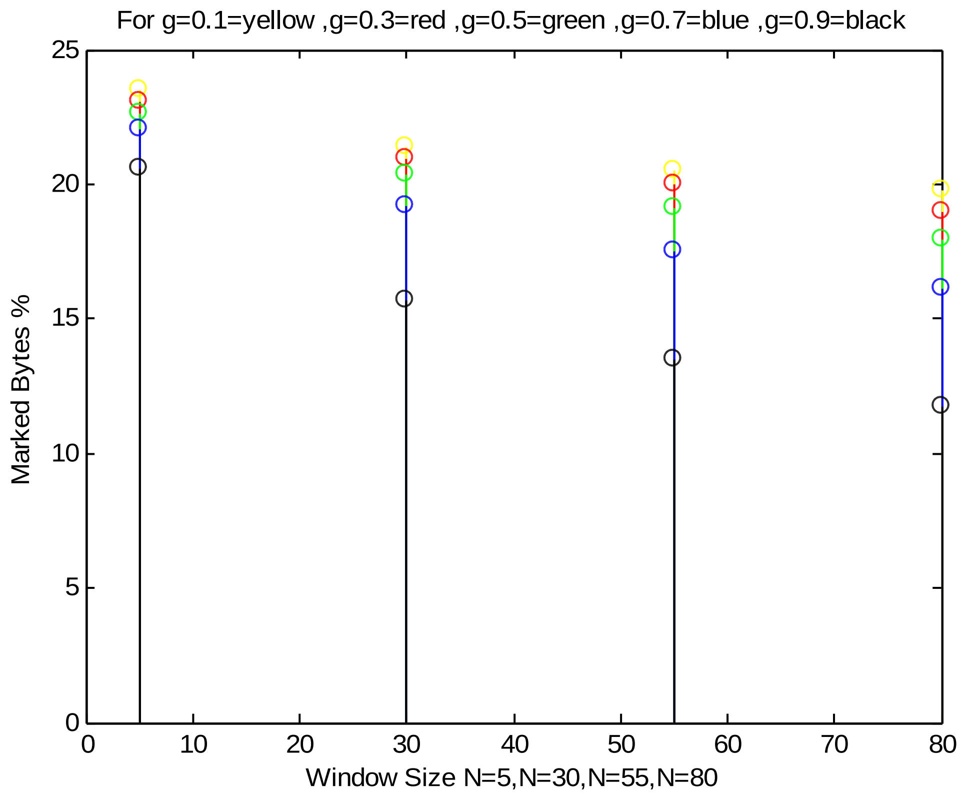

Figure 43. Terminator, B2, QP28. Marked traffic with the use of the EWMA mechanism.

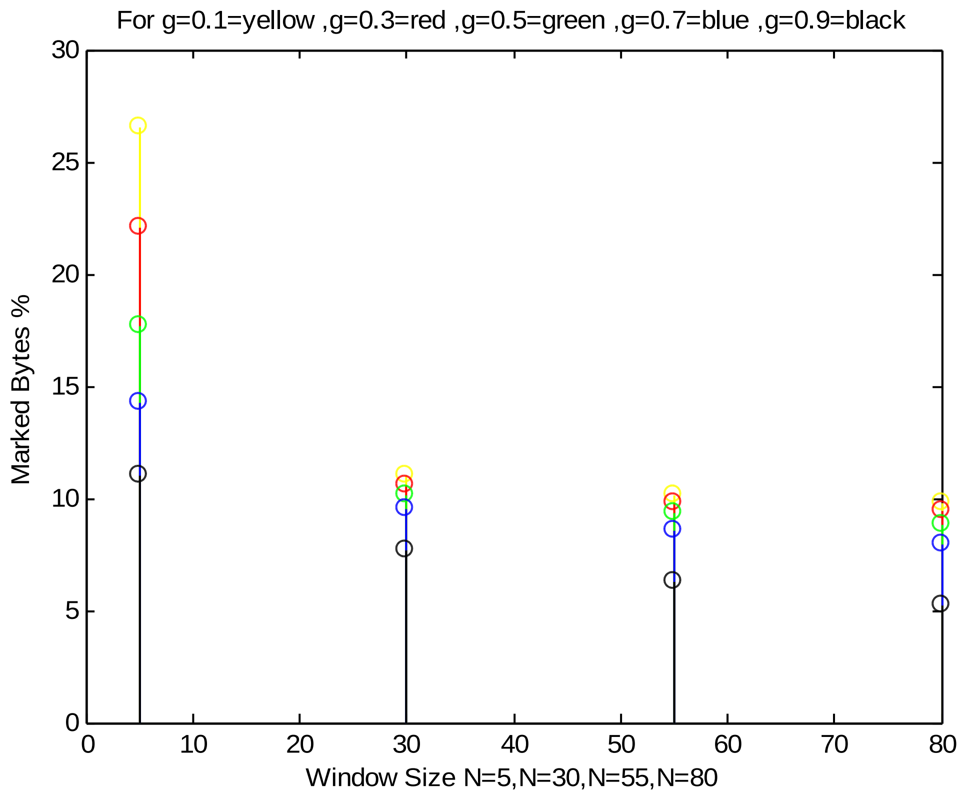

Figure 44. Kaet's Horizon, QP28. Marked traffic with the use of the EWMA mechanism.

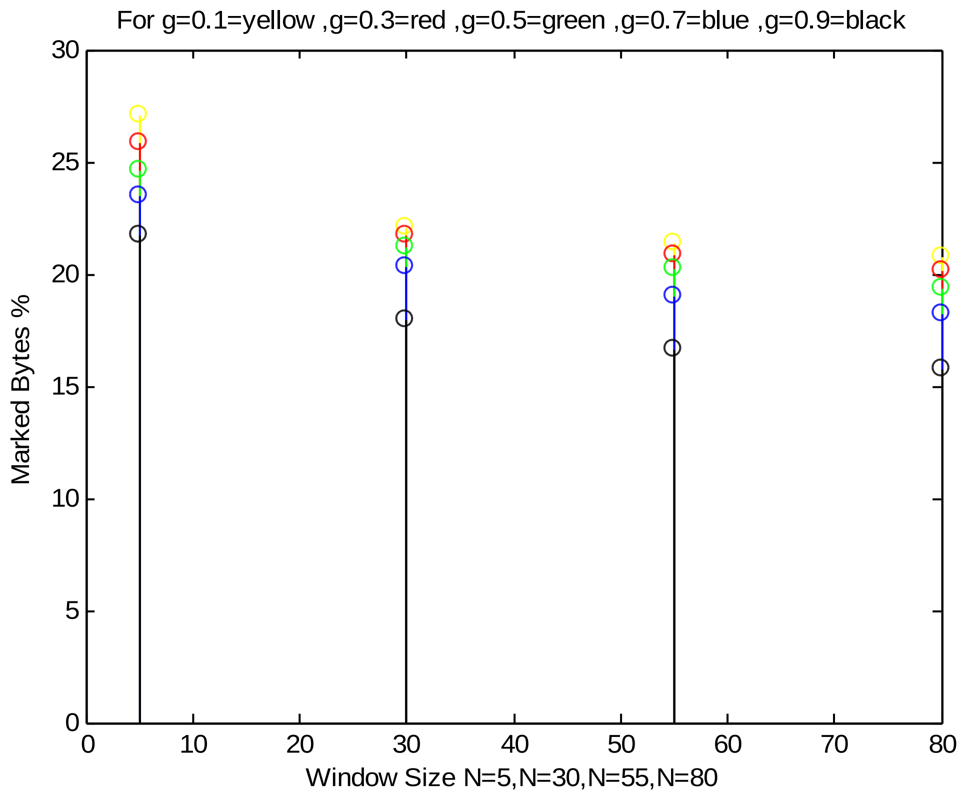

Figure 45. Kaet's from Mars to China, QP28. Marked traffic with the use of the EWMA mechanism.

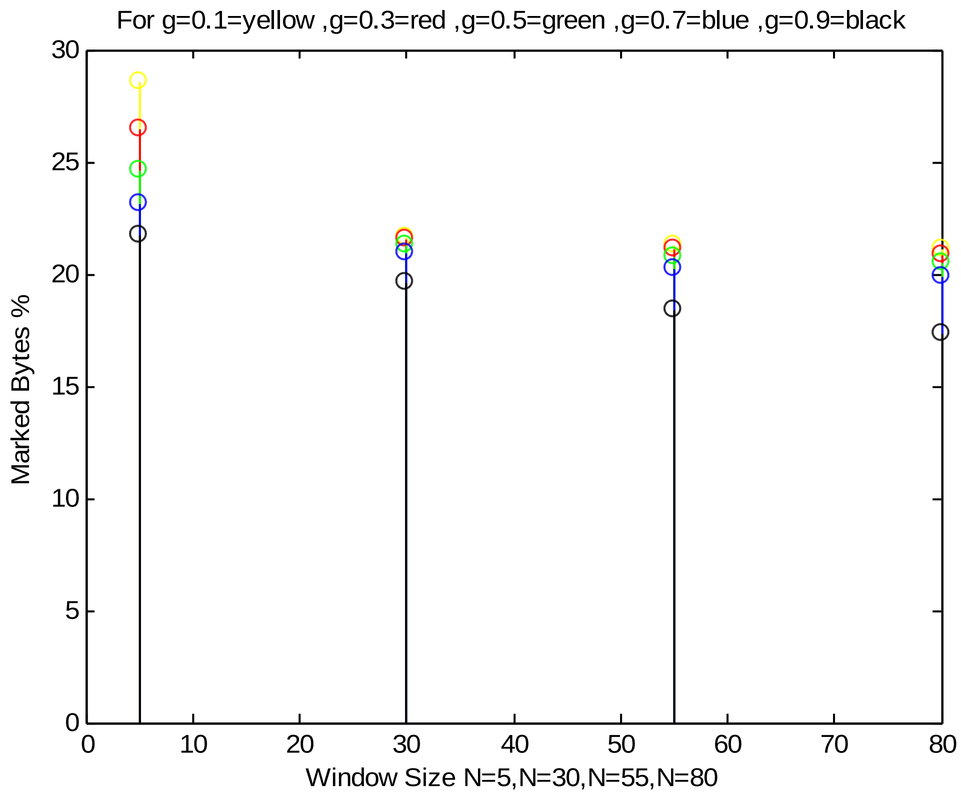

Figure 46. SonyDemo2, QP28. Marked traffic with the use of the EWMA mechanism.

4.5. Results with the use of GBTB

By studying the results of all 5 mechanisms (TB, LB, JW, MW, EWMA) presented in the previous Sections, it is clear that the TB and EWMA mechanisms achieve the best results in terms of not "penalizing" conforming users with marking their traffic as much as the other mechanisms do. This is intuitively expected, as the LB is actually a stricter version of the TB, and the JW and MW mechanisms do not possess the inherent advantages of the TB and EWMA mechanisms, i.e., to allow for larger transmission volumes from users who have been transmitting at less than their mean rate for a while (TB) and to be very lenient with the proper use of g and window size values (EWMA). Of course, the stricter nature of LB, JW and MW makes them more efficient in the case of non-conforming users.

One indicative example of the better performance of TB and EWMA are the results for the "Horizon" trace, where the TB can achieve (for a large bucket size) a marked traffic percentage of 6%, EWMA can achieve (for a large window size and a large g value) a marked traffic percentage of 5.2%, whereas the respective percentages for the LB, JW, MW mechanisms are all around 10%. Because of: a) the smaller complexity of the TB mechanism, in comparison with EWMA, and b) the fact that between the two, TB is stricter against non-conforming traffic than EWMA with large g and window size values, and c) the rather poor performance of all 5 mechanisms in the case of conforming but bursty video users, we were motivated to base our new mechanism, GBTB, on the Token Bucket mechanism. GBTB was presented in Section 3.6.

Table 4 presents the results in terms of the percentage of marked traffic of every trace, when applying the GBTB mechanism. It also presents again, as in Table 2, the results for the standard TB mechanism, when a very large (5 times the peak) bucket size is used, as well as the results regarding the required token generation rate in order to achieve a 1% and a 0.01% percentage of marked traffic. By exploiting the GoP pattern of each trace, GBTB is shown to largely outperform the standard TB mechanism, despite the fact that the bucket size used in GBTB is equal to the peak (i.e., 5 times smaller than the bucket used for TB). GBTB also outperforms TB in terms of the increase in token generation rate in order to achieve low percentages of marked traffic for conforming users. Compared to the mean rate of its trace, the standard TB mechanism needs, on average, a 1.6 times larger token generation rate than GBTB to achieve 1% marked traffic, and a 3.8 times larger token generation rate than GBTB to achieve 0.01% marked traffic. In many cases, as shown in the Table, the increase over the mean token generation rate with the use of GBTB needs to be very small in order to offer high QoS to conforming users. Therefore, with only this minor tradeoff in terms of its leniency, GBTB is a better candidate for video traffic policing than the classic mechanisms under study.

Movie	Codec	QP	GBTB	GBTB X times the mean value (for 1%)	GBTB X times the mean value (for 0.01%)	TB 5 x Bucket	TB X times the mean value (for 1%)	TB X times the mean value (for 0.01%)
Tokyo	B3	16	6.9	1.19	1.24	18.9	2.4	3.5
		28	10.2	1.33	1.40	23.3	3.4	5.7
		38	8.8	1.32	1.42	20.2	4.1	6.2
		48	6.7	1.23	1.31	15.5	2.3	3.9
	B7	16	7.5	1.19	1.26	19.8	2.5	3.6
		28	10.3	1.30	1.40	24.3	3.6	5.7
		38	8.2	1.32	1.42	20.6	3.0	5.9
		48	7.8	1.25	1.33	15.9	2.4	4.0
Silence of the Lambs	B3	16	19.7	2.96	3.26	27.5	6.1	9.3
		28	20.5	2.96	3.27	26.4	8.9	14.5
		38	17.3	2.67	3.33	21.6	7.0	14.4
		48	10.9	1.79	3.52	13.7	2.7	7.5
	B7	16	20.1	2.80	3.01	28.3	6.3	9.2
		28	20.6	2.86	3.16	27.0	9.4	14.6
		38	17.3	2.72	3.38	21.9	7.4	14.6
		48	10.9	1.90	2.60	13.3	2.8	7.7
Star Wars	B3	16	6.2	1.48	1.68	19.3	2.7	5.6
		28	6.7	1.44	1.65	19.2	3.2	8.4
		38	7.3	1.48	1.58	16.9	2.9	9.3
		48	5.9	1.31	1.56	11.4	2.1	7.0
	B7	16	6.5	1.39	1.55	20.2	2.7	5.7
		28	6.9	1.42	1.55	20.1	3.3	8.7
		38	7.6	1.42	1.53	17.4	3.1	10.5
		48	6.4	1.31	1.59	11.4	2.2	6.8
SonyDemo	B3	16	8.6	1.10	1.13	19.6	1.9	2.7
		28	11.1	1.17	1.23	17.6	2.0	3.4
		38	11.7	1.18	1.22	15.4	1.9	3.4
		48	9.4	1.13	1.16	14.3	1.7	2.3
	B7	16	9.2	1.10	1.18	20.9	1.9	2.6
		28	10.8	1.18	1.23	18.4	2.0	2.7
		38	11.6	1.17	1.21	15.8	1.9	2.5
		48	9.6	1.13	1.16	14.8	1.8	2.7
Nbc News	B3	16	1.5	1.02	1.05	11.9	1.7	2.3
		28	2.2	1.02	1.77	16.6	2.6	3.9
		38	2.8	1.03	2.98	11.9	2.3	4.6
		48	2.1	1.02	3.97	7.6	1.8	3.7
	B7	16	1.5	1.01	1.06	12.0	1.7	2.3
		28	2.2	1.02	1.50	17.1	2.6	3.9
		38	3.0	1.03	2.50	12.3	2.5	4.7

	48	2.3	1.02	3.59	7.9	1.9	3.9
Terminator	28	9.1	1.28	1.47	17.1	2.1	3.7
	38	7.2	1.30	1.71	15.1	2.1	2.9
	48	5.4	1.29	1.67	13.6	2.4	3.7
From Mars to China		2.2	1.02	1.04	17.2	2.2	4.0
Horizon		2.6	1.91	3.76	6.0	2.1	4.3
SonyDemo2	28	10.2	1.22	1.28	18.9	2.0	3.1
	38	10.8	1.19	1.24	16.4	1.8	3.8
	48	10.1	1.14	1.18	15.7	1.9	3.4

Table 4. Marked traffic (%) for various bucket sizes and token generation rates, with the use of the TB and GBTB mechanisms.

Finally, although this work is focused on the case of conforming video users, we also studied the case of a malevolent user entering the network and attempting to transmit at much higher mean and peak rates than the declared ones. More specifically, we studied the case where a user declares the traffic statistics of the Star Wars IV B7, QP28 trace, while in reality the used transmits the Tokyo Olympics B7, QP28 trace. The results for the TB, JW, MW, EWMA and GBTB mechanisms were, respectively:

- For TB, with a token generation rate equal to the mean and a bucket size equal to the peak, the percentage of marked traffic was 53.4%.
- For TB, with a token generation rate equal to the mean and a bucket size equal to 5 times the peak, the percentage of marked traffic was 51.9%.
- For JW, with a window size equal to 5, the percentage of marked traffic was 56.7%.
- For JW, with a window size equal to 80, the percentage of marked traffic was 53.3%.
- For MW, with a window size equal to 5, the percentage of marked traffic was 58%.
- For MW, with a window size equal to 80, the percentage of marked traffic was 54%.
- For EWMA, with g=0.1 and a window size equal to 5, the percentage of marked traffic was 56.4%. When increasing the window size to 80, the percentage of marked traffic dropped to 53.2%.

- For EWMA, with g=0.9 and a window size equal to 5, the percentage of marked traffic was 53.8%. When increasing the window size to 80, the percentage of marked traffic dropped to 51%.
- For GBTB, the percentage of marked traffic was 50.9%.

The above results show that our mechanism provides the least strict policing against non-conforming video users. However, the percentage of marked traffic with the use of GBTB is still very high, and the difference in its efficiency compared to the other mechanisms is small, especially when compared with their more lenient versions, since their stricter versions provide quite disappointing QoS for conforming users.

5. Conclusions

We have implemented and evaluated the efficiency of a number of classic traffic policing mechanisms on H.264 video users. Our results have shown that all the mechanisms used in our study are not capable of ensuring the satisfaction of the high QoS requirements of conforming video users, because of the burstiness of H.264-encoded video.

Hence, we used our results to specify which of the classic mechanisms performs relatively better (i.e., with less strictness for the conforming users), and after finding out that this was the token bucket mechanism we based our new mechanism, which is especially tailored for video traffic, on the token bucket. However, instead of using the classic approach, we exploited the periodic nature of the video frames' arrival, i.e., the GoP pattern of each video trace. With the use of 3 token buckets, one for each type of video frame, our mechanism GBTB is shown to clearly outperform all the classic mechanisms against which it is compared, in terms of providing high QoS to conforming video users, with only a minor tradeoff in terms of its leniency, both for conforming and non-conforming video users.

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 CommunicationsNetworks (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/hd/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, Glasgow, U.K., 2007, 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.