Video Traffic Modeling using Kolmogorov Smirnov Analysis in Broadband Network

Murizah Kassim[1], Rafidah Samsuri[1], Mat Ikram Yusof[1], Ruhani Ab Rahman[1], Naimah Mat Isa[1], Roslina Mohamad[1], Mahamod Ismail[2]

[1]*Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 UiTM Shah Alam, Selangor, Malaysia.*

[2]*Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia 43600 UKM Bangi, Selangor, Malaysia.*

murizah@salam.uitm.edu.my

***Abstract*—Video Traffic utilization is one of the major issues for Quality of Service (QoS) for network traffic especially in broadband network. Most network administrators are looking at providing best QoS and reliable traffic performances especially on video traffic. Analysis on recent trend and modeling video traffic activity is a crucial task in providing better bandwidth usage. This research presents an analysis on video network traffic in a Broadband Network in Malaysia. Real data from a telecommunications service company based for Business and Home network are collected. Traffic characterization is analyzed and new traffic parameters and model are presented. Goodness of fit (GoF) and Kolmogorov Smirnov (KS) test is used to fit the real traffic in getting the best Traffic distribution model. Results present four top video used in the network traffic which are You Tube, MPEG, TV on Streamyx and Dailymotion using standard video protocol. Fitted traffics presents Pareto model is best fitted on video traffic. Generalized Pareto (GP) with Empirical Cumulative Distribution function (CDF) distribution is identified as the best distribution model. The fitted Generalized Pareto model was identified based on lower Kolmogorov-Smirnov (KS) value and higher probability value (p-value). Test statistics for four particular distribution results at 5% level significance. GP characterization presents three important parameters which are shape, scale and location. A new mathematical formulation is derived based on control parameters gathered for future rate limiting algorithms.**

***Index Terms*—Video Traffic Modeling; Quality of Service; Kolmogorov Smirnov Analysis; Broadband Network.**

I. Introduction

Rapid increase of network traffic access in social networking media and video streaming has created complex set of data network traffic. Whenever network continues to operate, traffic are still continues to process larger than before and more bandwidth is utilized. Analyzing on traffic data especially on video traffic provides information such as average load, bandwidth used, Quality of Service (QoS), application pass through in different protocols and others network traffic entities [1, 2]. An increasingly of bandwidth usage would impact network capabilities of traffic performance in many networks [3, 4]. Thus, it is important to monitor bandwidth consumption activities in real time broadband network and analyze network traffic to allow better QoS [5]. Network traffic performances are unpredictable and have variety of traffic such as video, audio, photo, large and small data transfer and so many more [6].

Traffic models are used to estimate parameters and evaluate congestions of traffic performance. Many network traffic models has been identified however, real traffic models keep changing over time. Each traffic model has different characteristics to fits network environment distributions [7, 8]. Thus understanding of latest traffic model is important especially in a broadband network. Traffic model is design to verify the real broadband network performance. Performance technique of traffic modeling is to determine the best technique use for traffic congestion control [9]. There have three principle consideration parts in order to design traffic model which are need to ensure accurate and fit traffic modeling, able to decide on parameter setting include mean and peak bit rate requirements and able to evaluate the expected performance under a wide variety of traffic condition [10]. This principle was applied in multimedia communication, digital video broadcasting and video on demand services [11]. Real video traffic model is an important factor to design tele-traffic algorithm to control traffic in a network. One example of traffic control called rate limiting to constraint the congestion traffic by limiting amount traffic pass through in a network. Rate limiting on burst traffic is an enable to regulate the amount of bandwidth specific traffic uses on specific interface by limiting the amount of data receives or forwards for high speed network traffic [12].

This paper present real time video traffic characterization in identify real data collected at a broadband network. Data were studied which were collected at one telecommunication's center in Malaysia. Statistical properties on traffic throughput collection are analyzed. Best few statistical models of probability distributions and estimation parameters are used such as Maximum Likelihood Estimators (MLE) and goodness of fit (GoF) test. In this paper, statistical distribution graphs is tested using Kolmogorov Smirnov (KS) test and important variable estimation parameter is identified to get the best fitting model Kolmogorov Smirnov (KS) statistic. It is identified that KS is more powerful on discriminator and loses the invalid distribution when parameters are estimated [13].

KS test is a robust test that concerned only about the relative distribution of the data, D. When the numbers differ significantly, KS will aim to allocate numbers to test results which call p-values report. KS test will reject the null hypothesis if p is small.

II. Reviews On Video Traffic Model

Observed on broadband traffic data in today's scenario has presented multiple network application need to be controlled to keep traffic flow in low network management cost. Bandwidth management is one an important part to control traffic congestion especially in multimedia application, such as video on demand and video conferencing which exists various in multimedia resources traffic characteristics. Identify video traffic characteristics are important to realize the new traffic trend and parameters and to find optimal values of bandwidth requirement [14]. Previous research presents various performance model techniques are used to determine effective way for congestion control in broadband network such as modeling of MPEG video sequence. This predicted model was used mixture-exponential distribution by wavelet decomposition method to capture closely characteristics of real video sequence [15]. Every traffic models has different characteristics in correlation structure and marginal distribution in order to predict network performance. Several probability function distribution are identified in traffic model such as Normal, Lognormal, Weibull and Pareto distribution which are most applications of statistical methods to describe distribution of science and engineering [16]. One target of statistical analysis is to determine estimation parameters or sample statistic. Goodness of fit (GoF) test is one of the tools used for assessing the principal distribution and provides estimation parameters of real live traffic data [17]. In order to fit distribution function, several tests is applied such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Chi-Squared (CS) test to provide important statistical parameters of the fitted model [18]. Weibull model is also known as one of the methods to measure QoS [19]. Few comparison parameters are shown in table 1 for analyzing video traffic model and development of algorithms used in previous research [20-22]. Observed that, a number of techniques for evaluating the distribution are presented, including a numerical integration scheme and a number of parameters approximations, which retain the original form of the traffic modeling distribution.

III. Methodology

A. Data Collection Diagram

Traffic on a Broadband network is measured based on amount and type of traffic flow for 7 days in the network. Traffic data was collect by Policy Traffic Switch (PTS) and it was taken from two different customers set that are Streamyx business and Streamyx home users. Figure 2 shows that Policy Traffic Switch (PTS) are located at layer3 where it is in between Multi Service Edge (MSE) and Broadband Remote Access Server router in network system. PTS obtained data traffic from 3 ports which provides internet speed 62.5MByte each port. Network traffic data for 7 days with 3 hours interval time is analyzed. Network implemented configurations parameters as in Table 2.

Table 1
Comparison parameters between this research and others researcher

Research	Rate Limiting Algorithm on Video Traffic with Pareto Traffic Model in Broadband Network	A Multiscale Model for MPEG-4 Varied Bit Rate Video Traffic	Policing method for Markovian traffic descriptors of VBR MPEG sources over ATM networks	A Streaming Video Traffic Model for the Mobile Access Network
Network	Broadband Network	Movie videos	Asynchronous Transfer Mode (ATM)	Radio Network
Traffic	Video Traffic	Frame traffic and Group of picture (GOP) traffic	VBR MPEG video traffic	Frame traffic
Traffic Volume	Inbound and Outbound	Inbound	Inbound	Inbound and Outbound
Model	Pareto and Generalized Pareto Distribution	Gaussian and Symmetric Pareto	Markovian traffic model	Normal Distribution
Parameter	Shape, Scale and location	Shape and scale	-	Shape and scale
Schedule	Video Traffic Protocol	Small and large time scales	Cell rate	Frame size (64kbps and 32kbps)
Algorithm	Policing and Shaping Scheduling	Multilevel Queuing Scheduling	Policing	-
Result/ Performance	To optimize bandwidth utilization	To easily stored and transport of bit information.	To fully bandwidth utilization	The bitrates and codec settings is suitable for mobile access

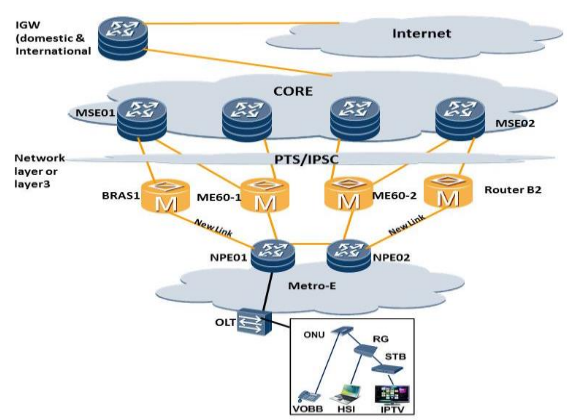

Figure 2: Policy Traffic Switch in Broadband Network

Table 2
Broadband network traffic parameter

Parameter	Value
Speed Limit, P_k	30 Gbps
Inter- Arrival Time	3 hours
Captured Time	00:00 to 21:00
Number of days	7 days
Minimum Speed	998.43 Byte
Maximum Speed	122.16KByte
Number of Data, Xi	560
Number of Protocol	10
Analytical Parameter	Shape, scale, location

B. Analysis and Future Development Flow

The research comprises two important phases which first, find fitted distribution then provide mathematical formulation model from the identified parameters as derived in Figure 3. Video traffic is collected and fits to Goodness of Fit (GoF) test. Kolmogorov Smirnov (KS) statistic test is used to compare the fit of distributions and decide the best fitted distribution based on Empirical Cumulative Distribution Function (CDF). GoF and KS test produced estimate parameters to represent characteristics of network traffic. The identification parameters are used to develop mathematical formulation for rate limiting burst video traffic controller algorithms. Statistical analysis on parameters of each collective data was studied and compared. Statistical value with lower of Kolmogorov-Smirnov (KS) and having higher probability value (p-value) is chosen as best traffic model.

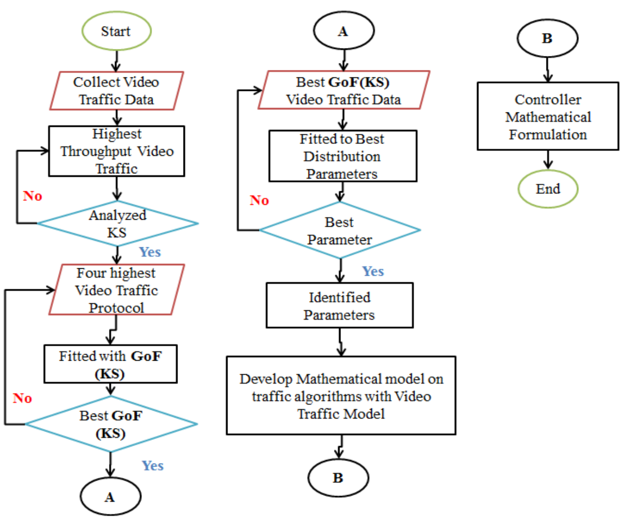

Figure 3: Flowchart of rate limiting algorithm

IV. Result And Analysis

A. Best Fitted Model

Video traffic data is characterized and fitted parameters model are presented. Figure 5 shows broadband video traffic that has presented Pareto model as best model for all multiple protocols which are You Tube, HTTP, Facebook, MPEG, TV over Streamyx, UDP, SSL, Google play, iTunes App Store and Dailymotion. Graph presents that 20% of critical points come from You Tube, HTTP, Facebook and MPEG while the rest 80% were minor points. This means the four particular protocols could make network traffic congestions.

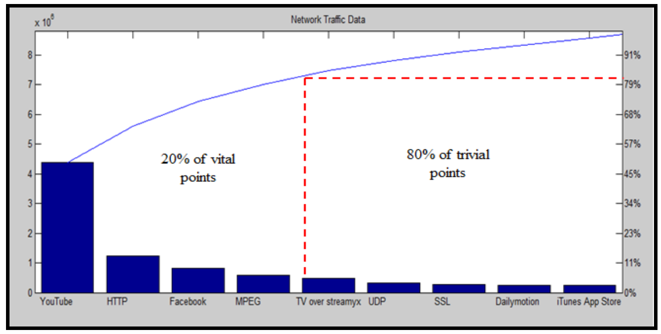

Figure 5: Data plot in Matlab

Figure 6 shows You Tube, MPEG, TV over Streamyx and Dailymotion are four data of commercial video services which are plotted in bar graph according to time and date for 7 days. It is observed that high bandwidth consumes comes from You Tube significantly huge compared to MPEG, TV over Streamyx and Dailymotion. Based on the daily traffic cumulative data, half of the traffic congestion are coming from YouTube streaming. YouTube as the main hitter contribute almost 50% of all traffic congestion. Majority of users prefer to stream during night time. Data shown an increase of Kbyte for streaming during AM vs. PM is almost double when throughput distribution is between 0-250Kbyte during day time compared to 0-450Kbyte during night time.

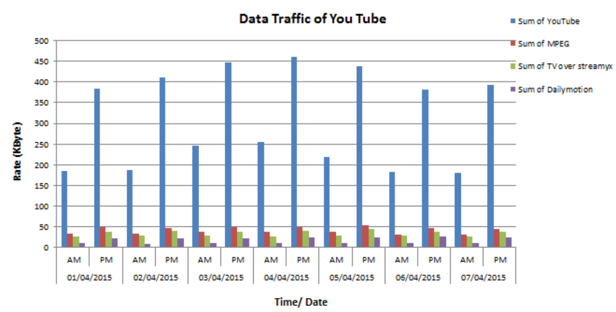

Figure 6: Time series of most high Video Traffic in day and night

Figure 7 shows best fitted traffic model on video traffic collected. Four best fitted distributions are derived which are Normal, Lognormal, Weibull and Pareto traffic model for You Tube, MPEG, TV over Streamyx and Dailymotion video traffic. Test statistics for four particular distribution results at 5% level significance are presented as in Table 3. Observed that each distribution have its own Kolmogorov-Smirnov (KS) statistic value and probability value (P-value). P-value for KS test is based on two hypotheses as follow:

H_0: Data follow the specified distribution
H_1: Data do not follow the specified distribution

Null hypothesis H_0 is rejected if P-value is lower than significance level (α) which is in this case is 0.05. Meanings data analysis does not follow the specified distribution. While if P-value is higher than significance level, Null hypothesis H_0 is accepted and concludes that data analysis is suitable for the specified distribution. Cumulative distribution function (CDF) statistic value in Generalized Pareto (GP) distribution for You Tube, MPEG, TV over Streamyx and Dailymotion is smallest than Normal, Lognormal and Weibull distribution which average is 0.09 close to real video traffic data. P-value in GP distribution is highest than other distribution model considered here which average 0.7 where the null hypothesis H_0 is accepted. In conclude Generalized Pareto (GP) distribution model tend to be more fits than Normal, Lognormal and Weibull distribution.

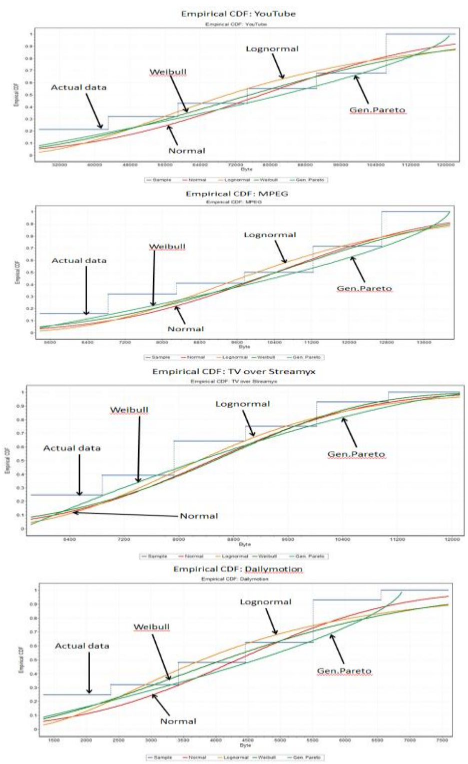

Figure 7: Empirical CDF of (a) You Tube (b) MPEG (c) TV over Streamyx (d) Dailymotion with different sample of distribution

B. *Parameters Characterization*

Parameters estimation result of Goodness of Fit (GoF) test for four types distribution model are presented in Table 4. All results obtained from the distributions below are based on two parameters which are scale and location parameters (σ and μ) and shape and scale parameters (α and β), except for Generalized Pareto, another additional criteria was added. Thus, Generalized Pareto distribution will give three reading of data which is B, σ and μ value to determine the shape, scale and location respectively. Location here refers to bandwidth traffic threshold. Statistical analysis of parameter estimation is to provide information measurement for distribution model validation and substantiate purposes.

Parameters estimation value of fitted Generalized Pareto (GP) distribution is to be continued used in analytical data for You Tube video traffic. Bandwidth can be varies linearly with B value. The higher the B value is, the greater the bandwidth would be consumes as in Figure 8.

Table 4
Parameters of resulting models

Distribution	You Tube	MPEG	TV over Streamyx	Dailymotion
Normal	σ=31681.0 μ=78048.0	σ=2788.6 μ=10432.0	σ=1797.4 μ=8475.6	σ=1918.1 μ=4332.2
Log normal	σ=0.49026 μ=11.16	σ=0.29253 μ=9.2127	σ=0.21105 μ=9.0228	σ=0.5657 μ=8.2392
Weibull	α=2.284 β=88537.0	α=3.8965 β=11472.0	α=5.3954 β=9117.2	α=1.9399 β=4965.9
Gen. Pareto	B=−1.4657 σ=15504.5 μ=15166.0	B=−1.5153 σ=14154.0 μ=4804.9	B=−0.70229 σ=4754.7 μ=5682.5	B=−1.4963 σ=9521.4 μ=518.01

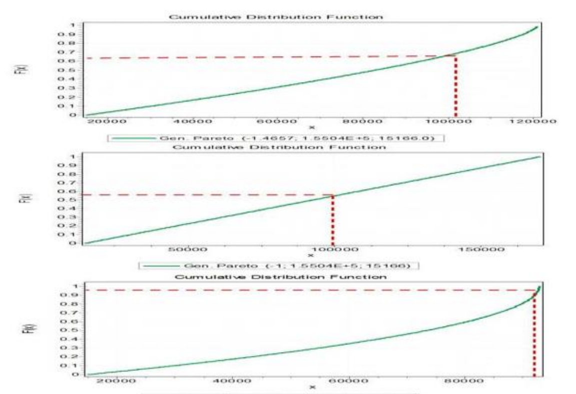

Figure 8: Empirical CDF of variable shape data with (a) B= -1.4657 (b) B=-1 (c) B= -2

Table 3
Goodness of Fit Test with Kolmogorov-Smirnov

Distribution	YouTube		MPEG		TV over Streamyx		Dailymotion	
	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Normal	0.14883	0.15103	0.12776	0.29438	0.11357	0.43351	0.1582	0.10854
Lognormal	0.15211	0.13485	0.16686	0.07854	0.12472	0.32128	0.18085	0.04487
Weibull	0.13541	0.23391	0.11698	0.39702	0.10595	0.52144	0.14456	0.17441
Generalized Pareto	0.09546		0.06477	0.96086	0.08764	0.74985	0.11248	0.4456

Based on Table 5, the optimum B value is -1.0000 where 55% of 100 Kbyte used. Thus, this can reduce the size of bandwidth consumption.

Table 5
Summary of variable shape parameter

Variable shape, B	Threshold	Result
-1.4657	100KByte	60%
-1.0000	100Kbyte	55%
-2.0000	100Kbyte	90%

V. CONCLUSION

Large scale network users such as business and home streamyx, monitoring for bandwidth activity is very important where the flow network usage is continue to increase in time. This research has successfully presents an analysis on video broadband network traffic. Traffic is analyzed and statistical analysis of fitted on best Empirical Cumulative Distribution Function (CDF) model is presented. Pareto distribution model is identified the best traffic video model based on four collected high video traffic which are YouTube, MPEG, TV over Streamyx and Dailymotion. Analyzed results also shows majority users prefer to stream during night time. Traffic congestion can be seen increasing rapidly during public holiday and weekend. The video traffic then is compared with Normal, Lognormal and Weibull distribution. From statistical method analysis, Generalized Pareto Model is identified as the best distribution model. The fitted Generalized Pareto model was identified based on lower Kolmogorov-Smirnov (KS) value and higher probability value (p-value) where conclude that null hypothesis is accepted. Generalized Pareto parameter characterizations are identified which are the shape, scale and location parameter. The parameters will be used as video traffic model in formulating a new mathematical model on rate limiting algorithms in future development. This traffic analysis is also valuable in controlling network traffic especially for bandwidth management or quality of service (QoS) in future traffic algorithms such as queue, scheduling or others tele-traffic algorithms.

ACKNOWLEDGEMENT

Authors would like to thank Universiti Teknologi MARA for support grant number 600-RMI/DANA 5/3/LESTARI (57/2015) for this research.

REFERENCES

[1] Kumaresh D, Suhas N, Krishna SGG, Anand S, Hegde M. SeaMoX: A Seamless Mobility Management Scheme for Real-Time Multimedia Traffic Over Cellular Networks. Advances in Signal Processing and Intelligent Recognition Systems. Springer International Publishing. (2016). p. 699-710.

[2] Varghese JM, Hariharan B, Uma G, Kumar R, editors. Adaptive Video Quality Throttling Based on Network Bandwidth for Virtual Classroom Systems. Second International Conference on Computer and Communication Technologies. Springer. (2016).

[3] Yamada Y, Miyao Y, editors. Impact of inactivity timer on performance of control plane in LTE core network under burst traffic. 2015 IEEE International Conference Communications (ICC). (2015).

[4] Bai Y, Hong F. Performance Impact of Wireless Mesh Networks with Mining Traffic Patterns. Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) (2008).

[5] Mohd Idris MF, Yusof MI, Azmat FH, Md Zain Z, Rahman RA, Kassim M. Broadband Internet performance (QoS measurement) view from home access gateway. IEEE 5th Control and System Graduate Research Colloquium (ICSGRC). (2014).

[6] Aktas MF, Haldeman G, Parashar M. Scheduling and flexible control of bandwidth and in-transit services for end-to-end application workflows. Future Generation Computer Systems. (2016). 56:284-94.

[7] Kassim M, Ismail M, Yusof MI. Statistical analysis and modeling of internet traffic IP-based network for tele-traffic engineering. ARPN Journal of Engineering and Applied Sciences. (2015).10(3):1505-12.

[8] Oliveira TP, Barbar JS, Soares AS. Computer network traffic prediction: a comparison between traditional and deep learning neural networks. International Journal of Big Data Intelligence. (2016).3(1):28-37.

[9] Min D, Loguinov D, Radha H. Statistical analysis and distortion modeling of MPEG-4 FGS. International Conference on Image Processing, (ICIP). (2003).

[10] Lambadaris, I., Devetsikiotis, M., Kaye, A.R., Ismail, M.R., Sharon, C.M., Fang, Y. and Huang, C. Traffic modeling and design methodologies for broadband networks. Canadian Journal of Electrical and Computer Engineering. (1995). 20(3):104-15.

[11] Richardson IEG, Riley MJ. Video quality of service in broadband networks. International Broadcasting Convention, IBC 95. (1995).

[12] Casoni M, Grazia CA, Klapez M, Patriciello N. QRM: A queue rate management for fairness and TCP flooding protection in mission-critical networks. Computer Networks. (2015). 93, Part 1:54-65.

[13] O'Hagan A, Murphy TB, Gormley IC, McNicholas PD, Karlis D. Clustering with the multivariate normal inverse Gaussian distribution. Computational Statistics and Data Analysis. (2016). 93:18-30.

[14] Pontes R, Coelho R. Admission Control for Video Traffic Streams with Scaling Characteristics. Journal of Communication and Information Systems. (2015).17(2).

[15] Yan H, Tao J, Yumin F, editors. A Novel Modeling of MPEG-4 Video Sequence. 8th International Conference on Signal Processing. (2006).

[16] Jie L, Aurelius A, Manxing D, Hantao W, Arvidsson A, Kihl M. Youtube traffic content analysis in the perspective of clip category and duration. Fourth International Conference on Network of the Future (NOF). (2013).

[17] Porter JE, Coleman JW, Moore AH. Modified KS, AD and C vM Tests for the Pareto distribution with unknown location and scale parameters. IEEE Reliability Transactions. (1992). 41(1):112-7.

[18] Yap-Peng T, Nagamani J, Hong L. Modified Kolmogorov-Smirnov metric for shot boundary detection. Electronics Letters. (2003) 39(18):1313-5.

[19] Kassim M, Ismail M, Yusof MI. Adaptive throughput policy algorithm with weibull traffic model for campus IP-based network. Information Technology Journal. (2014). 13(17):2632-44.

[20] Huang X-d, Zhou Y-h, Zhang R-f. A multi-scale model for MPEG-4 varied bit rate video traffic. IEEE Transactions on Broadcasting. (2004). 50(3):323-34.

[21] Yoo S-J, Kim S-D. Policing method for Markovian traffic descriptors of VBR MPEG sources over ATM networks. IEE Proceedings in Communications. (2001).148(2):70-6.

[22] Nyberg H, Johansson C, Olin B, editors. A streaming video traffic model for the mobile access network. IEEE VTS 54th Vehicular Technology Conference, VTC. (2001).