Peak Time Bandwidth Control Algorithms with Fitted Traffic Model on Broadband Network Youtube Video Traffic

Aini binti Azmi
Faculty of Electrical Engineering
Universiti Teknologi MARA UiTM Shah Alam
Selangor, Malaysia
iniamiza@gmail.com

***Abstract*— This paper presents an analysis of video traffic and fitted to best distribution traffic model to control bandwidth usage in a broadband network. The study scope comprised of collections of inbound YouTube video traffic for 7 days with the time-interval of each day is 3 hours. The broadband network is supported at 10Gbps line speed to Wide Area Network (WAN). The objective of this research is to characterize YouTube video traffic on broadband network, to fit the original traffic to best traffic model and Bandwidth control algorithm is developed based on the real live traffic and using the fitted traffic model. Results presents traffic characterizations are identified based on two (2) parameters Cumulative Distribution Functions (CDF) traffic model. Maximum Likelihood Estimator (MLE) technique is used to fit the best distribution model. Four best traffic model is identified which are Extreme Value, Weibull, Normal and Rician traffic model. Among the four, Weibull shown as the best fitted model that presents value of MLE=-1178.4 with the Scale α=9.49411e+08 and Shape β=2.81324. Bandwidth Control Algorithms is developed based on Peak Time of day and night. Performance shows the bandwidth controlled as bandwidth save, processing time, bucket capacity and cost. Research benefits in the development of design network especially for bandwidth used on Video used in a network with the result enhanced the performance for 54%.**

Keywords—video traffic, cumulative distribution function, traffic model, weibull, extreme value, normal, rician

I. Introduction

Internet played a very important role in this era. The operation of internet started at 1960's when the development of ARPANET, which would become the first network to use Internet Protocol. Nowadays, most of activities in this world need to use internet to complete each of task or to communicate among each other either for a short distance or long distance communication. There are many applications build from the use of internet such as social networking sites, online streaming sites and the latest is cloud computing. The application of video traffic is widely use either for individual, company or commercial such as Youtube, television websites or individual websites which provide the free drama online streaming. The use of video traffic usually involved high speed of bandwidth data to avoid buffering and delayed to the user. Because the number of user are increasing year by year, the demand to supply the internet bandwidth are getting high. Internet Service Provider (ISP) need to find way to allocate the bandwidth to the customer as per requested to avoid traffic delay or bad service happen [1]. The application of video traffic consists of two component which are sound and image[2]. Both components will vary base on the length of the video itself. In this situation, the long the length of the video, the more bandwidth will be use and the process of sending the video or uploading the video to the internet will be longer. Here is one of the applications by the user which will cause to high bandwidth utilization. In order to control the bandwidth usage, there are many modeling and algorithms that can be applied in the video traffic network to make sure customer satisfied with the service. Traffic shaping is a process of controlling the bandwidth by passing the burst data to the available bucket data below the committed rate with a bit time of delay. It can be applied at the interface of the networks to manage the incoming and outgoing traffic. Traffic shaping can be use to avoid saturation from happening and it can control and manage network latency[1].Therefore, the Demand Side Management algorithm was purposed to reduce the delay and enhanced the performance of the service.

The research of behavior of network traffic have been done and it shows that traffic policing and shaping can only decrease the data rates of the network but the aggregate traffic behavior is not affected[3]. In other way, the bandwidth usage can be control by identifying the model of the distribution which suite to the video traffic data. Some of the research present that the modeling traffic can reduce the traffic burst and at the same time can control the bandwidth use in customer side[2], [4]. It is use the Maximum Likelihood Estimator (MLE) technique to find the best distribution model to be use by fitted the real data to Cumulative Distribution Function (CDF) using MATLAB software[5]. It is also compared four type of distribution to get the best method and the value of the parameter in the distribution is used to control the network traffic flow. There are also some other test being used to compare the distribution such as the application of Goodness-of-Fit statistical toolkit which can produce the best algorithm in the internet traffic_which will not be use in this research[6]. There are many different research have been done to control the traffic in the network and it is depends on the component that involved such as the traffic analysis on the feature like packet size, frequency, inter-arrival time, traffic burst and others[7][8][9]. Generalize Extreme Value model being used to model the extreme events by using the block maxima approach. Weibull distribution model was applied to capture the transformation of inter-arrival process and the parameter being used to zoom in from session to flow and to packet level inter-arrivals[8]. Some research use the parameter value of α (scale) to control their traffic but in this research, the value of β (shape) being used.

This paper presents a statistical analysis on video traffic network and best fitted distribution for YouTube data. It is supported 10Gbps Committed Access rate speed line. Throughput video traffic is collected for 7 days with 56 numbers of tracers in GByte. The inter-arrival time for each tracer is 3 hours. The throughputs are fitted to CDF using MATLAB software and the values of MLE log are taken. Best four distributions from the four highest MLE log value are selected which are Extreme Value, Weibull, Normal and Rician. Weibull distribution is identified as a best distribution model with 2 parameters which one of the parameter will be used to control the bandwidth in the traffic data. This research is benefits in identifications of tele-traffic engineering in modeling video traffic distribution.

II. REVIEW ON MODELING TRAFFIC AND CONTROL ALGORTIHMS

There are many control algorithms being used in video traffic network. It is depends on the subject that the research is focus on such as to reduce the burst, to control the bit rate, delay, packet receive and others. One of the researches was done in controlling the delay by using Leaky-Bucket based rate control algorithm. This algorithm can reduce the burst in the video traffic and the video can be transmitted smoothly without buffer[10]. In other research, it is focusing on the packet size on the video streaming in LTE network. The novel scheduling algorithm is presented to give the better service in LTE network by applying the algorithm at the enhancement layer. This algorithm can improve the packet delivery ratio and serve the better service[11]. Data burst always happened when there is heavy flow of traffic occurred in the network. To overcome this situation, Bit Rate Throttling algorithm was proposed to reduce the burst in RTP which is use to transfer the real time data in the network. This algorithm is focusing to control the burst at the video encoder by separated the RTP data into packet before sending to the destination. This algorithm can avoid the data spike occurred and can reduce buffer in the video traffic network [12] .

The performance in networking is very important to make sure the service can be delivering in a good condition. In order to complete this performance, the traffic model should be apply with the best defined parameters in each distribution model. Several distribution model has been identified that can be used to measure the traffic.

A. *Extreme Value Distribution*

Extreme Value distribution is a 3-parameter distribution. It is use parameter μ as location, σ as scale and ξ as shape (tail). Extreme value usually being use in finance, structural engineering and earth sciences. The mathematical equation for this distribution as shown in Equation (1).

$$F(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

B. *Normal Distribution*

Normal distribution is a most distribution used in the application. It is sometimes called bell curve because of the shape of the graph is like a bell shape. Normal distributions are important in statistics to represent real-valued random variables whose distributions are not known. There are three parameters in this distribution which are dataset,x,(location) mean, μ and (scale) standard deviation, σ. The mathematical equation as in Equation (2).

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

C. *Weibull Distribution*

Weibull distribution consists of two parameters which are scale, α and shape, β. The mathematical equation for cumulative distribution function has shown in Equation (3). For application, the shape parameter is an important part as a control measurement for the algorithm by increasing or decreasing the value. Weibull distribution can be use in wireless communication.

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

D. *Rician Distribution*

Rician distribution used two type of continuous parameters which are $v \geq 0$ and $\sigma > 0$ with the dataset, x. It is also consist of Q_1, which is Marcum Q-function. This distribution usually being used in microwave application which called Rician Fading Distribution. The mathematical equation as shown in Equation (4).

$$F(x) = 1 - Q_1\left(\frac{v}{\sigma}, \frac{x}{\sigma}\right)$$

III. METHODOLOGY

A set of video traffic data was collected in 7 days. The inter-arrival time for each data was measured in 3 hours each. It is containing of 4 times in a day and 4 times in a night with the total number of tracers are 56 for a week. To characterize the data, the statistical analysis has to be performed by comparing the best fitted cumulative distribution function (CDF) with the highest Maximum Likelihood Estimator (MLE) value. Then the real data will go through the policing process by setting up the value of Committed Access Rate (CAR) for day and night with the different values to control the receive bandwidth. The policing process will cut off the access bandwidth which called 'burst'. Normally, burst will happened in video traffic because of the overload of data flows in the network. After policing process, the shaping process will perform to fill the burst in available bucket to prevent the byte lost but it will cause a bit delay in processing time in this step. When all process completed, then the selected distribution model with the highest value of MLE is applied. The parameter from the selected model is used to control the bandwidth by adjusting its value. The results will be discussed in the next session.

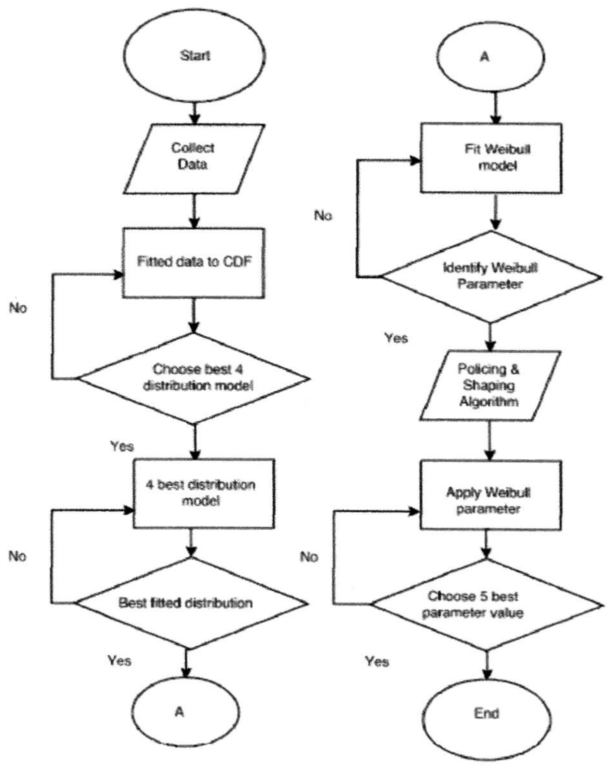

Figure-1. Flowchart of the Research

IV. ANALYSIS AND RESULT

Video data collected of 7 days with 3 hours interval is analyzed.

A. *Statistical Analysis*

Figure-2 present the throughput from the real data which collected in 7 days with the total number of tracers is 56 tracers. The data is taken in GByte. The graph shows a heavy throughput in day time which is from 7am to 6pm while for night time from 7pm to 5am, the reading of the throughput is lesser than day time.

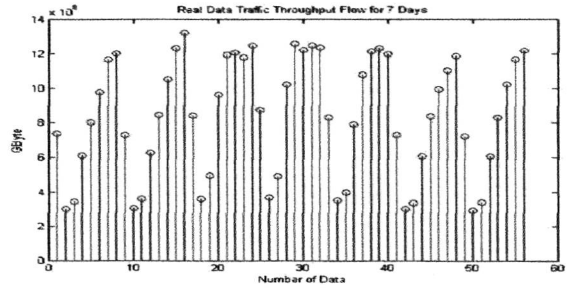

Figure-2.Real live data collected for 7 days

Figure-3 shows the selected distribution model for the highest four type of distribution that fitted to CDF base on the measurement on MLE log value. The analysis for the graph has shown in Table-1 with the highest MLE log value is -1177.56 for Extreme Value Distribution. But in this condition, Extreme Value distribution will not be use because it is use 3-parameters in the mathematical equation. So Weibull will be a selected distribution model with the MLE log value -1178.4 with the parameter value of α (shape)= 9.49411e+08 and β (scale)= 2.81324.

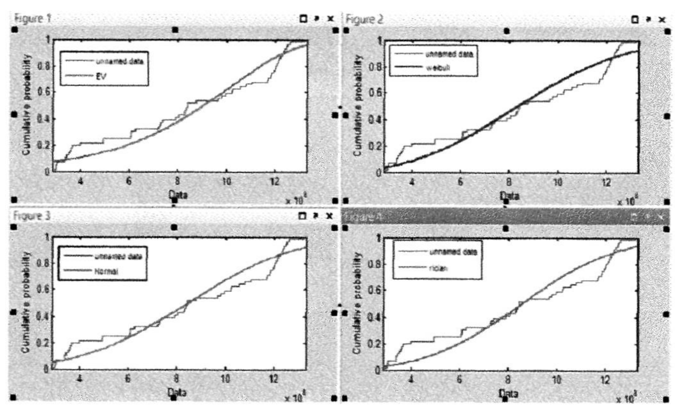

Figure-3. CDF fit on best MLE throughput

Table-1: Best MLE for Youtube Traffic

Traffic Model	MLE	Parameter	
Extreme Value	-1177.56	μ	σ
		8.42914e+08	2.76779e+08
Weibull	-1178.4	α	β
		9.49411e+08	2.81324
Normal	-1179.4	μ	σ
		8.42514e+08	3.42156e+08
Rician	-1179.84	s	σ
		7.88109e+08	3.19663e+08

B. *Control algorithms with Policing and Shaping*

Figure-4 presents the throughput from real data before policing and the throughput after policing for day time only. The green line in Committed Rate Level and Burst Throughput graph from Figure-3 shows a 1Gbps rate bandwidth which represents a threshold guide. The red color of the data in the graph shows the burst traffic exist in the network. In policing, all the burst will be cut off and the bandwidth will not be use anymore. Even though the bandwidth can be saved, this process will cause to byte loss which customer will not receive the full packet of data. Figure-5 presents the bandwidth control algorithms called policing on Youtube traffic for day and night with the different value of threshold which the rate for day is 1Gbps and for night is 0.5Gbps. The threshold for night is reduced because the bandwidth usage in night time is lesser than day time.

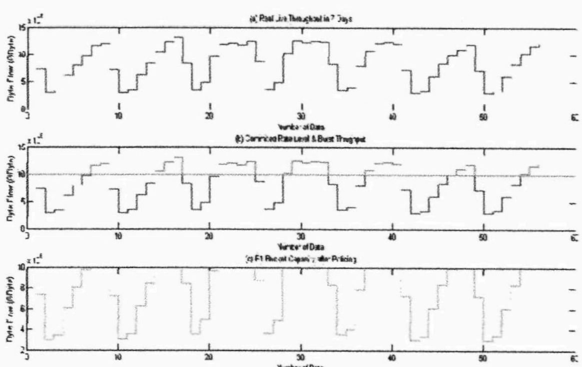

Figure-4.Policing on day time

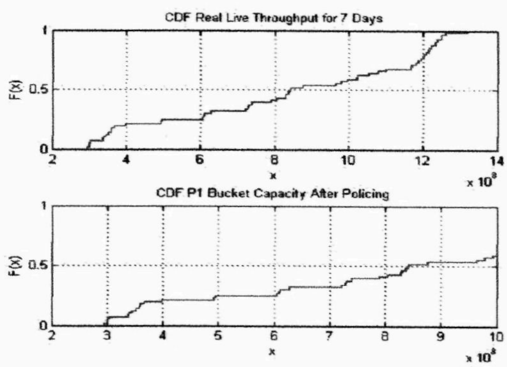

Figure-5.CDF before and after policing for day time

Figure-5 presents the CDF graph for day time before and after applying the policing algorithm. The graph after policing shown that the reducing of the throughput value in the traffic. It is because the burst from the data already being cut off.

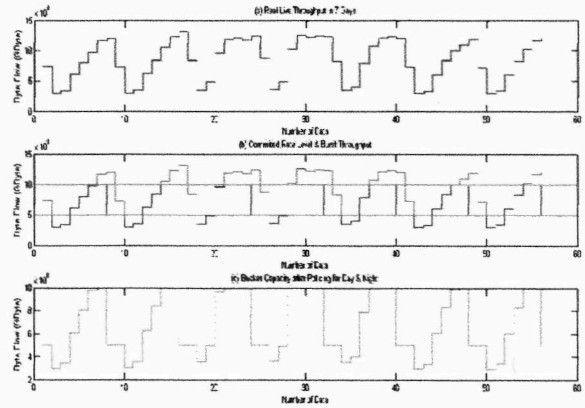

Figure-6.Policing for day and night

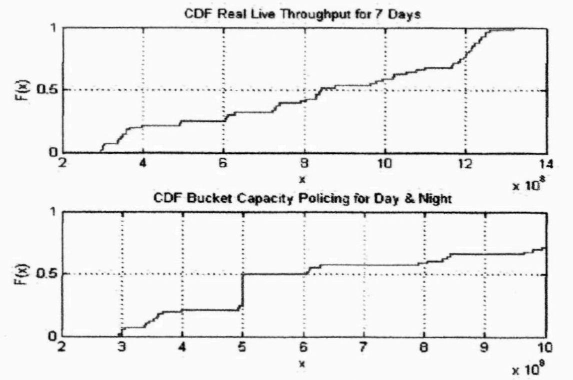

Figure-7.CDF before and after policing for day and night

In Figure-7, the policing was applied separately for day and night. This condition can be referred in the CDF graph for policing day and night which there is an increasing value of F(x) at the point of x=5, which is the threshold for night time is 0.5Gbps. The readings of the policing graph also being reduce because of the cut off burst in the traffic.

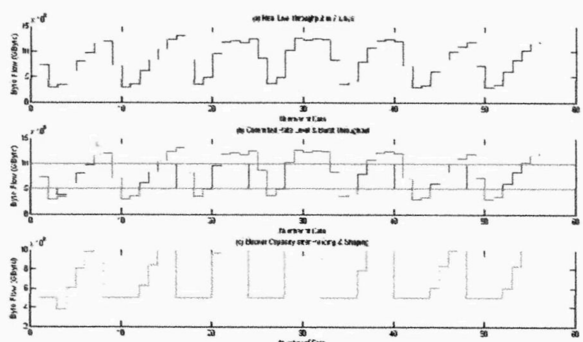

Figure-8.Policing and shaping algorithm for day and night

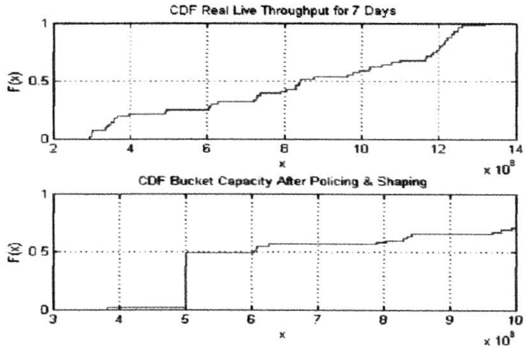

Figure-9.CDF graph before and after policing and shaping algorithm for day and night

To overcome the byte lost in the policing process, shaping algorithm is applied in this research. Shaping process will passing the burst data to the available bucket data below the committed rate with a bit time of delay without wasting the packet data. Customer will received the full packet of data without fail even they will face with a bit delayed. This can be proved by referring to the Figure-8, which is the bucket data in Committed Rate Level and Burst Throughput graph shown that all the burst in the network was passing to the next available bucket to save the bandwidth. The CDF graph in Figure-9 also shown the reading which near to consistent value after applying the policing and shaping algorithm. The analysis of the policing and shaping algorithm from the real data is shown in Table-2. The processing time can be reduced and the bandwidth also can be saving in a high value. So the higher the bandwidth can be saving, the speed of the traffic will increase and the delay and buffer can be reduced.

Table-2. Analysis on policing and shaping from the real data

Parameter	Value
Total bandwidth save	25548000
Total burst shape	34490000000
Byte lost	-8332300000
Total Process time (before)	86513
Total Process time (after)	38707
Different Process Time	47806
Minimum value	383380000
Maximum value	2586300000
Mean value	1458800000

C. *Application of Weibull Distribution Model*

From this research, we identified that Weibull distribution model is the best model fitted to the video traffic data given. Figure-9 to Figure-18 presents the different result of Weibull graph using the different value of shape parameter which is β in the Equation (3). This result is produced from 7 days data collected but if the research is done by day, the result might be differ for each day. Table-3 present the analysis fitted parameter estimation for CDF Weibull. From the table, the reading for bandwidth save, burst shape and byte lost will reduce if the β parameter is increase while the processing time will increase when the β also increase.

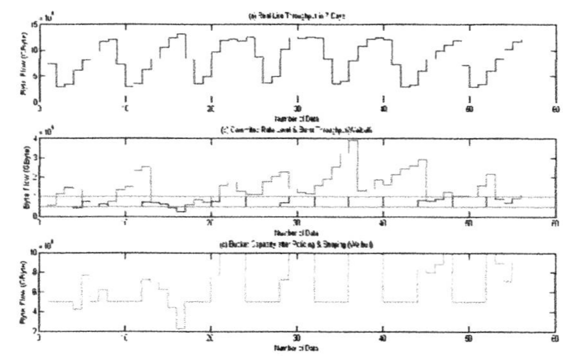

Figure-10.The application of Weibull distribution for parameter α=949411000 and β=2.81324

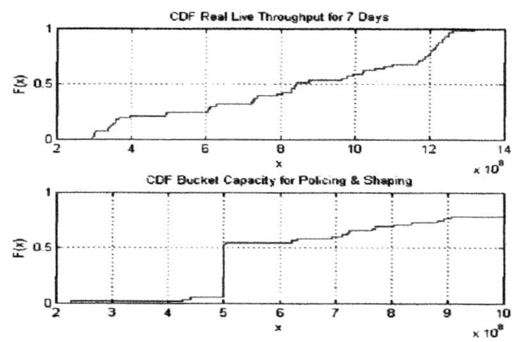

Figure-11.CDF graph of Weibull distribution for parameter α=949411000 and β=2.81324

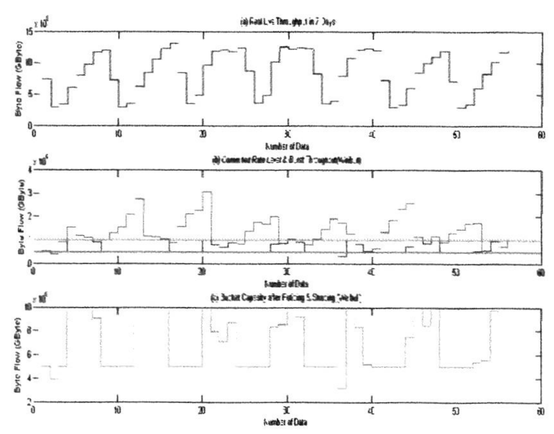

Figure-12.The application of Weibull distribution for parameter α=949411000 and β=4

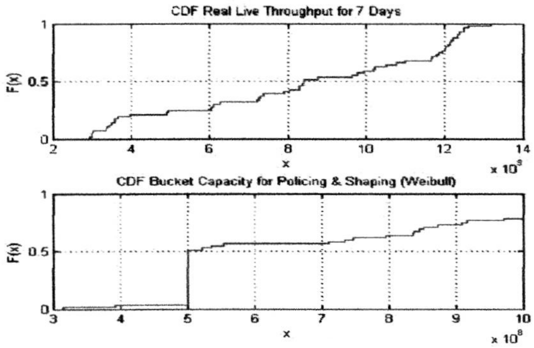

Figure-13.CDF graph of Weibull distribution for parameter α=949411000 and β=4

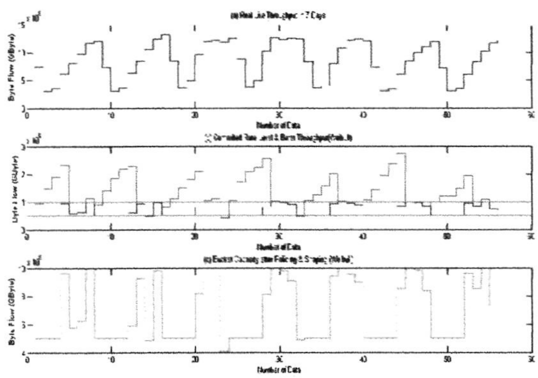

Figure-14.The application of Weibull distribution for parameter α=949411000 and β=6

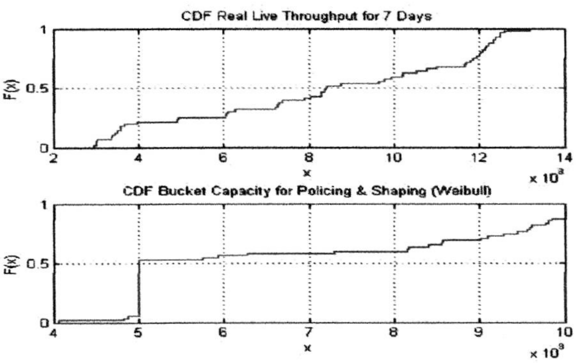

Figure-15.CDF graph of Weibull distribution for parameter α=949411000 and β=6

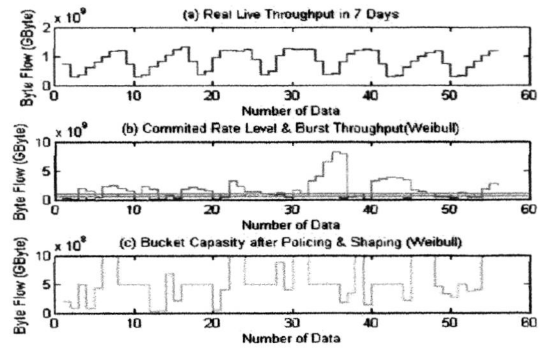

Figure-16.The application of Weibull distribution for parameter α=949411000 and β=1

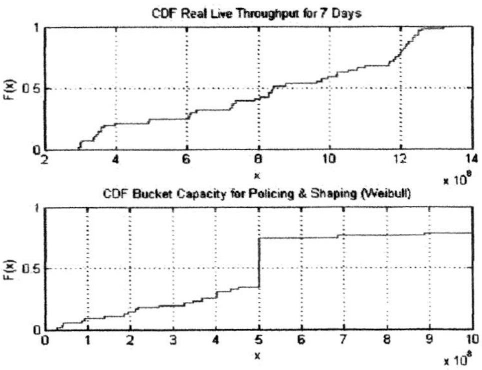

Figure-17.CDF graph of Weibull distribution for parameter α=949411000 and β=1

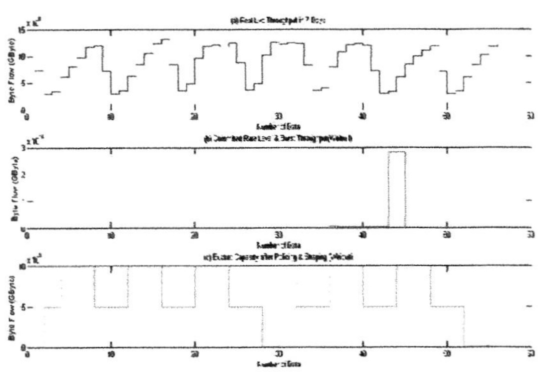

Figure-18.The application of Weibull distribution for parameter α=949411000 and β=0.1

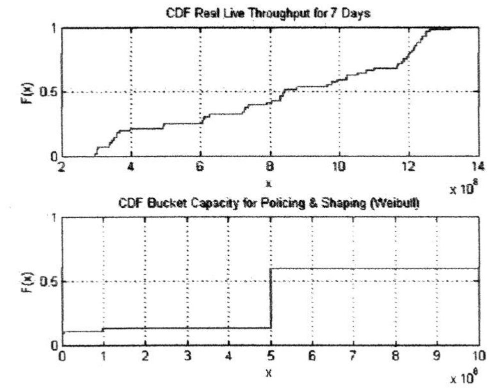

Figure-19.CDF graph of Weibull distribution for parameter α=949411000 and β=0.1

Table-3.Comparison of result using different β value in Weibull parameter

	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
	Algorithm policing Weibull (949411000, 2.81324)	Algorithm policing Weibull (949411000, 4)	Algorithm policing Weibull (949411000, 6)	Algorithm policing Weibull (949411000, 1)	Algorithm policing Weibull (949411000, 0.1)
Total bandwidth save (bps)	23532000	18845000	19249000	42318000	233010000000
Total burst shape	3.1768e+10	2.5441e+10	2.5987e+10	5.7130e+10	3.1457e+15
Byte lost	-1.0653e+10	2.5441e+10	2.5987e+10	5.7130e+10	3.1457e+15
Total Process Time (before)	80520	73937	76014	107810	4.5285e+09
Total Process Time (after)	36872	37590	37840	30771	35633
Minimum value	2.1697e+08	3.9290e+08	4.0640e+08	4.1563e+07	1.3022e-05
Maximum value	2.8903e+09	3.0653e+09	2.7425e+09	8.1316e+09	2.8141e+15
Mean value	1.3010e+09	1.3076e+09	1.3272e+09	1.8715+09	1.0684e+14

V. CONCLUSION

The analysis for video traffic data and the best modeling in fitted distribution is presented. The throughput for 7 days using YouTube data is collected and analyzed. Four best fitted distribution which are Extreme Value, Weibull, Normal and Rician are used. Weibull distribution is selected as the best video traffic distribution which produced the higher value of MLE log for the data. The traffic policing and shaping algorithm being applied to control the burst exist in the traffic. Then the parameter from the Weibull distribution being used and the result shows the improvement of the traffic in term of speed and bandwidth usage. This research is benefits in identifications of tele-traffic engineering in modeling video traffic distribution.

ACKNOWLEDGMENT

Author would like to thank to Dr. Murizah Kassim for the full support in doing this research.

REFERENCES

[1] A. M. Aladwani, A. Gawanmeh, and S. Nicolas, "A Demand Side Management Traffic Shaping and Scheduling Algorithm," *J. Comput. Sci.*, pp. 205–210, 2012.

[2] K. S. F. A. M. H. O. A. Chaudhry, "Statistical analysis of H . 264 video frame size distribution," *J. Comput. Sci.*, vol. 5, no. September 2010, pp. 1978–1986, 2011.

[3] D. Daniel-simion and G. Dan-horia, "Traffic shaping and traffic policing impacts on aggregate traffic behaviour in high speed networks," pp. 465–467, 2011.

[4] M. Kassim, M. Ismail, and M. I. Yusof, "STATISTICAL ANALYSIS AND MODELING OF INTERNET TRAFFIC IP- BASED NETWORK FOR TELE-TRAFFIC ENGINEERING," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 3, pp. 1505–1512, 2015.

[5] M. Kassim, M. A. Abdullah, and M. M. Sani, "INTERNET TRAFFIC ANALYSIS WITH GOODNESS OF FIT TEST ON," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 21, pp. 9892–9898, 2015.

[6] S. Guatelli, B. Mascialino, A. Pfeiffer, M. G. Pia, A. Ribon, and P. Viarengo, "Application of statistical methods for the comparison of data distributions," *J. Comput. Sci.*, vol. 00, no. C, pp. 2086–2090, 2004.

[7] F. Zhang, W. He, and X. Liu, "Defending Against Traffic Analysis in Wireless Networks Through Traffic Reshaping," *2011 31st Int. Conf. Distrib. Comput. Syst.*, 2011.

[8] M. A. Arfeen, K. Pawlikowski, D. Mcnickle, and A. Willig, "The Role of the Weibull Distribution in Internet Traffic Modeling," *J. Comput. Sci.*, 2013.

[9] A. Y. Dahab and H. Hasbullah, "Predicting Traffic Bursts using Extreme Value Theory," *2009 Int. Conf. Signal Acquis. Process.*, pp. 229–233, 2009.

[10] J. Bai, Q. Liao, X. Lin, and Y. J. Zhung, "A rate control algorithm for vbr video encoding and transmission," *J. Comput. Sci.*, pp. 0–3.

[11] Y. Lin, "An Novel Scheduling Algorithm for Video Stream in LTE," *2012 Sixth Int. Conf. Genet. Evol. Comput.*, 2012.

[12] M. Y. Modi and S. Kasula, "Bit Rate Throttling Algorithm on Video over RTP," *2013 Nirma Univ. Int. Conf. Eng.*, pp. 1–4, 2013.