Adaptive Policing and Shaping Algorithms on Inbound Traffic Using Generalized Pareto distribution

Nor Azura Bt Uyop@Ayop Faculty of Electrical Engineering Universiti Teknologi MARA UiTM Shah Alam azuraayop@gmail.com

Abstract— This paper present an analysis of inbound internet traffic and development of Adaptive Policing and Shaping Algorithms on inbound internet traffic and fitted to traffic model. The objective of this research is to characterize inbound internet traffic collected on real live IP-based campus network. Then, traffic is fitted to best traffic model and percentage level Policing and Shaping algorithm is developed to control the bandwidth used. The research scope is based on collected of internet traffic on IP-based network real live traffic at 16 Mbps speed line. Open Distribution Fitting application is fitted to the collected data to identifying the best distribution and the results presents Generalized Pareto shows the highest value for best fitted traffic model. Log likelihood estimation technique is used to fitted the best 2parameter CDF compared to Weibull, Normal and Rician distribution model. The percentage level 5% under original bandwidth used is developed on policing and shaping algorithms to control bandwidth used. Result present performances upgraded around 3% of time processing and approximately 73% of bandwidth saved. This result help to expand the view of new idea in modelling the tele-traffic algorithm based on bandwidth management and time processing improvement.

Keywords— internet traffic, bandwidth management, time processing, policing, shaping, algorithm, Generalized Pareto, Weibull, Normal, Rician, distribution

I. Introduction

The requirement to have improvement in internet traffic management nowadays became crucial. The users demand is increasing sky rocketed day by day. Through research and observation on daily life of internet traffic, researcher has created so many algorithm that has one sole mission, it is to handle internet traffic effectively [1]. The successful in emerging the next generation of telecommunication such as LTE cannot promise the best internet performance if they can't manage their shared wireless resources in the most efficient way [2]. Although there is a lot of development of algorithm, method or scheme are developed to control network traffic in an IP-based network but still, organizations faces high volume of traffic used every day [3]. Traffics that run in the IP-based network may comprise from different network protocols and heterogeneous applications which cause burst traffic with the used of new technology on the internet [4].

This paper present an analysis of inbound internet traffic and development of Adaptive Policing and Shaping Algorithms. The collected data traffic is fitted to best traffic model and percentage level Policing and Shaping algorithms is developed to control the bandwidth used. The research scope is based on collected of internet traffic on IP-based network real live traffic at 16 Mbps speed line. Open Distribution Fitting application is fitted to the collected data to identifying the best distribution and the results presents Generalized Pareto selected as the best fitted traffic model. The percentage level 5% from original bandwidth used is developed on policing and shaping algorithms to control bandwidth used. Result presented the performances for each algorithm. This result help to expand the view of new idea in modelling the tele-traffic algorithm based on bandwidth management and time processing improvement.

II. Literature Review

The good bandwidth management guarantees the performance and QoS running at their optimum level of services. The bandwidth management purpose is to search the finer levels of dynamic control to achieve the maximum efficiency, thus it need the admission control and traffic provisioning to make more bandwidth available to QoS traffic. If we could adjust the reserved bandwidth to handle the fluctuation of the incoming traffic rate, traffic can accept more service requests [5]. The burst data above the committed rate become the main concern for the Internet Service Provider (ISP). Thus, in order to suppress the burst traffic shaping is used to flatten these peaks and balance the demand on network resources over larger periods of time. One of the studies related to the shaping algorithm is Demand Side Management (DSM), where the process of implementing measures and standards on the customer side, mainly used to reduce electricity demands, energy intensity of Internet traffic and bandwidth utilization [6]. The concept is achieved by flattening the response of the network and reducing the peak times, while offering the same level of service to the consumer [7].

The other alternative ways to manage internet traffic is the policing algorithm. It is also one the frequent use method to handle the internet traffic. The traffic policer drops data packets from traffic flows that violate the maximum allowed data rate [8]. Policing algorithm can be implemented at various network protocols at different or more standard network layers in Open Systems Interconnection model (OSI) [9]. The policing algorithm can also preventing misuse of apportioned bandwidth in ATM systems [10].

Internet traffic modeling is also an important and essential to understand. It is one of the major mechanism that can contribute to solve performance-related issues of current and future networks. Many efforts have been focused on modeling of source traffic related to specific application-level protocols, also with the purpose to conduct realistic network traffic simulations and emulations (i.e. generating synthetic traffic in real networks) [11].

III. Methodology

Method includes the statistical analysis on collected data and algorithms Policing and Shaping and fitted traffic to traffic model.

A. Flowchart



Figure 1: Methodology Flow

Figure 1 shows flowchart of the method used in this research. The data given is support 16 Mbps committed rate and the daily traffic is perfectly meet the provider committed rate, below 16 Mbps for every received packet. Thus, to create

the 'burst situation' the committed rate was decreased to 15 Mbps. While, inbound traffic throughput is in Mbit and collected daily for seven days with 10 minute interval and 672 traces.

The analysis of the characteristic to determine the suitable distribution is using MATLAB Open Distribution Fitting application, 'dfittool', the maximum likelihood estimation (MLE) is determined. Using the fitted CDF, the distribution that suit with data and give highest value is selected as a model and for this paper Generalized Pareto model is selected. For the 15 Mbps speed and 10 minute time interval, the bucket calculated is 9000 Mbit and this value will be use as a maximum bucket limit. Policing Algorithm is first to analyze with the given data. With this algorithm all the burst data will be cut-off at the 9000 Mbit. Then, the result provided is combine with the second algorithm, the Adaptive Bandwidth Algorithm. The adaptive bandwidth algorithm is, where the data below the committed rate will be multiplied with 105%. Thus, the bandwidth and the data will have different of 5%. For this algorithm, the bandwidth size will be flexible with the amount of data at that meantime. The second algorithm is Shaping Algorithm. With this algorithm, the data will be cutoff and the remaining will be add to the next bucket. The loop of cut-off and adding will continuously running until it find the empty space and pour all the remaining data. The bucket limit is 9000 Mbit and the end result shall not exceed this value. The result provided is also combine with the Adaptive Bandwidth Algorithm. Continuing analysis to this data, the earlier selected distribution, Generalized Pareto was applied together with Policing Algorithm and Shaping Algorithm respectively. Then, the Generalized Pareto shape parameter, k was changed with a few value to identified the best parameter value. All the result attain was compared, observed and analyzed to determine the best application that can suit with the internet traffic characteristic.

B. Traffic Distribution Model

The four selected distribution in this analysis is reviewed.

a. Generalized Pareto Distribution

The Generalized Pareto distribution is a generalization of the Pareto distribution and often used in risk analysis. The Generalized Pareto distribution has a location parameter, θ , a scale parameter, σ which must be strictly greater than 0 and a shape parameter, k [12]. The CDF equation present as per Equation 1;

$$F(x) = \begin{cases} 1 - \left(1 + k \frac{(x-\mu)}{\sigma}\right)^{-1/k} & k \neq 0\\ 1 - exp\left(-\frac{(x-\mu)}{\sigma}\right) & k = 0 \end{cases}$$
(1)

b. Weibull Distribution

The Weibull distribution normally use to model failure time data. It is frequently used to analyses the reliability of system or condition due to its capability and flexibility in modelling many different type of data, based on the value of the shape parameter, β [13]. Various shapes of the Weibull distribution can be revealed by changing the scale parameter, α , and the shape parameter, β . The Weibull CDF are commonly represented in Equation 2:

$$F(x) = \mathbf{1} - exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$
(2)

The Weibull distribution is particularly versatile because it reduces to an exponential distribution when $\beta = 1$.

c. Normal Distribution

The Normal distribution is widely use in data statistic determination [14]. This distribution's pdf graph is symmetric with a bell shape. It is also called the "Gaussian curve" after the mathematician Karl Friedrich Gauss [15]. The commonly use to represent the statistical data for common event such as academic result distribution, human height distribution and also capable to characterize the complex data distribution such as internet traffic distribution. The mathematical explanation for Normal distribution is represented by Equation 3;

$$F(x) = \emptyset\left(\frac{x-\mu}{\sigma}\right) \tag{3}$$

Where μ is the mean and σ is standard deviation. These parameters show the characteristic of normal distribution.

d. Rician Distribution

The Rician distribution is commonly use to model scattered signal which have stronger line-of-sight that reach a receiver by multiple paths [16]. The Rician distribution has same derivation with Rayleigh distribution, but Rayleigh distribution is specialized model when there is no line-of-sight. The equation of Rician distribution CDF is given by Equation 4 below with Q1 is the Marcum Q-function.

$$F(x) = 1 - Q_1\left(\frac{v}{\sigma}, \frac{x}{\sigma}\right) \tag{4}$$

IV.Result

Result presents the traffic characterizations and traffic performance on bandwidth based on the develop algorithms.

A. Modelling traffic

QoS (quality of service) and performance in internet traffic shall be balance to meet the user expectation. Hence, it is a must to continuously searching the best traffic model that can give the optimum approached. The parameter of models defined must be related to the actual performance traffic measures which are to be predicted from the traffic model [15]. By using the statistical function in MATLAB software, the best four distribution model has been chosen, which provided the Maximum Likelihood Estimation (MLE) log-likelihood value and perfectly match with the desired measurement curve. From this result, the first objective for this research to characterize inbound internet traffic collected on real live IP-based campus network is achieved.

The graph from CDF fitting application is;



Figure 2: CDF graph of Generalized Pareto Distribution The graph from CDF fitting application is;





The graph from CDF fitting application is;







Figure 5: CDF graph of Rician Distribution

Table 1: Maximum likelihood value.

Distribution:	Generalized Pareto	Weibull	Normal	Rician
Log likelihood:	-5393.37	-6132.41	-6157.74	6164.24

Table 2: Parameter distribution value

Distribution	Generalized Pareto	Weibull	Normal	Rician
Parameters	-	-	-	-
shape	-5.35433	4.68312	-	-
scale	50158.2	8645.13	2310.25	2386.99
location	0	-	7905.020	-
non- centrality	-	-	-	7511.54

The result in Table 1 shown the MLE maximum log-likelihood value for four distribution model. As shown in Table 2, the Generalized Pareto was identified as the best traffic characterization with the scale parameter, $\sigma = 50158.2$, shape parameter, k = -5.35433 and threshold correction, $\theta = 0$. The shape parameter value was changed to three others value -1, -3 and -5.5 purposely, to see the different in the result of processing time, bandwidth save, packet loss and CDF graph curve. From the result the best value parameter will be decided.

B. Traffic Performance on Policing and Shaping

This section completed the second objective in this research, it is to develop Adaptive Policing and Shaping Algorithms with percentage level on Inbound Traffic based on traffic characterization.

1. Policing vs Shaping Comparison



Figure 6: Policing Algorithm.



Figure 7: CDF graph before and after application of Policing Algorithm.



Figure 8: Shaping Algorithm.



Figure 9: CDF graph before and after application of Shaping Algorithm.

Figure 6 and Figure 8 shown the data before, during and after the algorithm is applied. While the Figure 7 and Figure 9 present the CDF graph before and after the algorithm is applied. Both graph and CDF graph presented the characteristic of algorithm respectively. The Figure (g) obviously shown how the data accumulated to the next bucket and due to this characteristic the burst is overwhelm.

Table (3): Policing vs Shaping Algorithm.

	Algorithm policing	Algorithm policing and shaping	Difference
Total bandwidth save	102 500	1705 500	
(Mbps)	183.509	4705.500	4521.991
Total burst shape			
(Mbit)	110110.000	2823300.000	2713190.000
Packet loss (Mbit)	110110.000	23209.000	86901.000
Total Process time			
(before) (minute)	5902.400	9013.600	3111.200
Total Process time			
(after) (minute)	5780.100	5876.600	96.500
Different Process			
Time (minute)	122.339	3137.000	3014.661
minimum			
value (Mbit)	993.911	993.911	0.000

maximum value (Mbit)	9367.800	28691.000	19323.200
mean value (Mbit)	7905.000	12106.000	4201.000

From the Table (3), the total bandwidth for policing was saved about 183.509 Mbps and 4705.5 Mbps for Shaping Algorithm. The total burst shape gap between the two algorithms is too big, 2,713,190 Mbit. It is because in the data for shaping algorithm, is not discarded. All the burst data will be added to the next bucket, due to this characteristic the burst is overflow and the processing time became slower compare to the policing algorithm. The number of packet loss is also quite big, 86,901 Mbit. It is because of the policing algorithm totally cut-off all the burst data while the shaping algorithm was bring the data forward and there is less packet loss in the shaping algorithm concept.

C. Traffic Performance on Policing and Shaping using fitted traffic model

2. Policing with GPD (different k value)

By changing the value of shape parameter, k. the graph and the CDF characteristic can be observed in Condition 2, 3 and 4. Condition 2 and Condition 3 was set higher than the fitted shape parameter while for the Condition 4, the parameter was set a little bit lower 0.5 from the fitted shape parameter. The value cannot be lower than -5.5 or it will be unable to further the simulation. Parameter in Condition 1 was the given value during the fitted CDF evaluation.



Figure 10: Policing with GPD – Condition 1 (k = -5.35433).



Figure 11: CDF graph of Policing with GPD – Condition 1 (k = -5.35433).



Figure 12: Policing with GPD – Condition 2 (k = -1).



Figure 13: CDF graph of Policing with GPD – Condition 2 (k = -1).



Figure 14: Policing with GPD – Condition 3 (k = -3).



Figure 15: CDF graph of Policing with GPD – Condition 3 (k = -3).



Figure 16: Policing with \overline{GPD} – Condition 4 (k = -5.5).



Figure 17: CDF graph of Policing with GPD – Condition 4 (k = -5.5)

Table (4): Comparison of k (shape) parameter, Policing with GPD

	condition I	condition 2	condition 3	condition 4
	Algorithm policing gp (-5.35433, 50158.2,0)	Algorithm policing gp (- 1, 50158.2,0j	Algorithm policing gp (-3, 50158.2,0)	Algorithm policing gp (- 5.5, 50158.2,0)
Total bandwidth save (Mbps)	188.607	18486.000	4963.800	47.716
Total burst shape (Mbit)	113160.000	11092000.000	2978300.000	28630.000
Packet loss (Mbit)	113160.000	11092000.000	2978300.000	28630.000
Total Process time (before) (minute)	5908.300	18429.000	9357.000	5622.100
Total Process time (after) (minute)	5782.600	6104.300	6047.800	5590.300
Different Process Time (minute)	125.738	12324.000	3309.200	31.811
mininum value (Mhti)	155.410	70.493	16.578	27.604
maximum valuc (Mbit)	9367.800	50092.000	16719.000	9119.700
mean value (Mbit)	7912.900	24681.000	12532.000	7529.600

Comparison of the different value of shape parameter, k is nicely arranged in the Table (4). From this table the bandwidth, time processing and packet loss for each parameter can be differentiate from one another. Condition 1 is the given parameter from the CDF fitted evaluation. While Condition 2, 3 and 4 is the observation value of different shape parameter. Condition 2 shown the extreme burst shape, 11,092,000 Mbit while the Condition 4 shown burst shape much lower and considered as the best value among the others. Condition 4 also have the faster processing time and also much lower packet loss compared to the other conditions. It seems the Condition 4 shape parameter, k = -5.5 is the best value for this model.

3. Shaping with GPD (different k value)



Figure 18: Shaping with GPD – Condition 1 (k = -5.35433).



Figure 19: CDF graph of Shaping with GPD – Condition 1 (k = -5.35433).



Figure 20: Shaping with GPD – Condition 2 (k = -1).



Figure 21: CDF graph of Shaping with GPD – Condition 2 (k = -1).



Figure 22: Shaping with GPD – Condition 3 (k = -3).



Figure 23: CDF graph of Shaping with GPD – Condition 3 (k = -3).



Figure 24: Shaping with GPD – Condition 4 (k = -5.5).



Figure 25: CDF graph of Shaping with GPD – Condition 4 (k = -5.5).

Table (5):	Comparison	of k	(shape)	parameter,	Shaping	with
GPD						

,	condition 1	condition 2	condition 3	condition 4
	Algorithm shaping gps (- 5.35433,50158.2,0)	Algorithm shaping gps (- 1,50158.2,0)	Algorithm shaping gps (- 3,50158.2,0)	Algorithm shaping gps (- 5,5,50158.2,0)
Total bandwidth save (Mbps)	554 758	6138700.000	1274900.000	116.690
Total burst shape (Mbit)	332860.000	3683200000.000	764960000.000	70014.000
Packet loss (Mbit)	7101.200	720600.000	726830.000	-103740.000
Total Process time (before) (minute)	6280-100	4099200.000	856670.000	5864.900
Total Process time (after) (minute)	5910.300	6703.100	6710.000	5787.100
Different Process Time (minute)	369.839	4092500 000	849960-000	77.793
minimum value (Mbit)	309.993	2774.900	16719.000	43.814
maximum value (Mbit)	13979 000	11006000.000	2278900.000	9948.900
mean value (Mbit)	8413.300	5506300.000	1150700.000	7868.300

Table (5) shown the comparison of the different value of shape parameter, k for Shaping with Generalized Pareto Distribution application. From this table the bandwidth, time processing and packet loss for each parameter can be differentiate from one another. Condition 1 is the given parameter from the CDF fitted evaluation. While Condition 2, 3 and 4 is the observation value of different shape parameter. Condition 2 shown the extreme burst shape, 3,683,200,000 Mbit while the Condition 4 shown burst shape much lower and also have the faster processing time compared to the other conditions. It seems the Condition 4 shape parameter, k = -5.5 also is the best value for this model.

From the earlier analysis, the result of modelling data with generalized pareto distribution with shape parameter, k = -5.5 is considered the best value for this model. Therefore, in Table (4) the comparison between policing algorithm with and without model distribution will be analyze.

4. Comparison of GPD Policing Algorithm

Table (6): Policing vs Policing with GPD

9	Algorithm	Algorithm policing gp (-	
	policing	5.5,50158.2,0)	Difference
Total bandwidth save			
(Mbps)	183.509	47.716	-135.793
Total burst shape			-
(Mbit)	110110.000	28630.000	81480.000
Packet loss (Mbit)	110110.000	28630.000	81480.000
Total Process time			
(before) (minute)	5902.400	5622.100	-280.300
Total Process time			
(after) (minute)	5780.100	5590.300	-189.800
Different Process			
Time (minute)	122.339	31.811	-90.528
minimum			
value (Mbit)	993.911	27.604	-966.307
maximum			
value (Mbit)	9367.800	9119.700	-248.100
mean value (Mbit)	7905.000	7529.600	-375.400

If look at the bandwidth save value, the policing algorithm without model distribution save 183.509 Mbps compare to policing algorithm with model distribution 47.716 Mbps. But, the burst shape and processing time is much lower at policing algorithm with model distribution.

5. Comparison of GPD Shaping Algorithm

Table (7): Shaping vs Shaping with GPD

	Algorithm policing and shaping	Algorithm policing and shaping gps (- 5.5,50158.2,0)	Difference
Total bandwidth save			
(Mbps)	4705.500	116.690	-4588.811
Total burst shape (Mbit)	2823300.000	70014.000	2753286.000
Packet loss (Mbit)	23209.000	103740.000	80531.000
Total Process time (before) (minute)	9013.600	5864.900	-3148.700
Total Process time (after) (minute)	5876.600	5787.100	-89.500
Different Process Time (minute)	3137.000	77.793	-3059.207
minimum value (Mbit)	993.911	43.814	-950.097
maximum value (Mbit)	28691.000	<u>994</u> 8.900	-18742.100
mean value (Mbit)	12106.000	7868.300	-4237.700

The bandwidth save value for shaping algorithm without model distribution is higher, 4705.5 Mbps compare to shaping algorithm with model distribution 116.69 Mbps but, the burst shape is high, 2,823,300 Mbit. The packet loss is high for the shaping algorithm with model distribution, 103,740 Mbit. The shaping algorithm is mean to be not to loss packet, if the loss is too big, then it is not meet with the main purpose of this algorithm.

The comparison and discussion for Table (6) and Table (7) complete the last objective for this paper, it is to compare the policing and shaping performance on bandwidth used, processing time and packet loss.

V. Conclusion and future work

The best fitted modelling distribution for the collected data is presented. Analysis and observation of the result and provided graph is already been done. Considering the burst shape, packet loss and processing time value seem like the Policing Algorithm with Generalized Pareto Distribution application (shape parameter, k = -5.5) is the most reliable for the collected data. However, it doesn't mean that the other algorithm is not good and cannot be use. Depend on the purpose and medium of the telecommunication, the other algorithm maybe more effective. In future, this algorithm can be used to characterize the other data and may include the analyzation for adaptive bandwidth.

Reference

- [1] H. Awad, H. Ibrahim, S. M. Nor, and B. M. Khammas, "Ambiguity and Concepts in Real Time Online Internet Traffic Classification," International Journal of Engineering and Technology (IJET), vol. 6, no. 1, pp. 403–410, 2014.
- [2] B. Al-manthari, N. Nasser, N. A. Ali, and H. Hassanein, "Efficient Bandwidth Management in Broadband Wireless Access Systems Using CACbased Dynamic Pricing," 2008 33rd IEEE Conference on Local Computer Networks (LCN), pp. 484–491, 2008.
- [3] M. Kassim, M. Ismail, K. Jumari, and M. I. Yusof, "A Survey: Bandwidth Management in an IP Based Network," international Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 6, no. 2, pp. 168–175, 2012.
- [4] C. Labovitz, A. Ahuja, F. Jahanian, B. Ave, and A. Arbor, "Experimental Study of Internet Stability and Backbone Failures," Fault-Tolerant Computing, 1999. Digest of Papers. Twenty-Ninth Annual International Symposium, pp. 278 285, 1999.
- [5] H. Guo and G.-S. Kuo, "A Dynamic and Adaptive Bandwidth Management Scheme for QoS Support in Wireless Multimedia Networks," 2005 IEEE 61st Vehicular Technology Conference, vol. 3, pp. 2081 – 2085, 2005.
- [6] A. M. Aladwani, A. Gawanmeh, and S. Nicolas, "A Demand Side Management Traffic Shaping and Scheduling Algorithm," 2012 Sixth Asia Modelling Symposium, pp. 205–210, 2012.
- [7] S. Vilvert and R. Teive, "Demand Side Management For Residential Consumers By Using Direct Control On The Loads," Power System Management and Control, 17-19 April 2002 Conference Publication No. 488 0 IEE 2002, pp. 233–237, 2002.

- [8] S. Ghazal and J. Ben-Othman, "Traffic Policing Based on Token Bucket Mechanism for WiMAX Networks," Communications (ICC), 2010 IEEE International Conference, pp. 1 – 6, 2010.
- [9] M. Kassim, M. Ismail, and M. I. Yusof, "A New Adaptive Throughput Policy Algorithm On Campus IP-Based Network Internet Traffic," Journal of Theoretical and Applied Information Technology, vol. 71, no. 2, pp. 205–214, 2015.
- [10] B. Lague, C. Rosenberg, and F. Guillemin, "A Generalization Of Some Policing Mechanisms," INFOCOM '92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE, pp. 767–775, 1992.
- [11] A. Dainotti, A. Pescap, P. S. Rossi, G. lannello, F. Palmieri, and G. Ventre, "An HMM Approach to Internet Traffic Modeling," IEEE Globecom 2006, no. vii, pp. 1-6, 2006.
- [12] A. Y. Dahab, A. Said, and H. Bin Hasbullah, "Quality of Service Metrics using Generalized Pareto Distribution," 2010 International Symposium on Information Technology, vol. 2, pp. 822–825, 2010.
- [13] M. A. Arfeen, K. Pawlikowski, D. Mcnickle, and A. Willig, "The Role of the Weibull Distribution in Internet Traffic Modeling," Teletraffic Congress (ITC), 2013 25th International, pp. 1 8, 2013.
- [14] M. Zhang and Y. Zhou, "Approximate Calculation about Standard Normal Distribution with Genetic Programming," 2010 Third International Conference on Information and Computing, vol. 3, pp. 17–20, 2010.
- [15] M. Kassim, M. A. Abdullah, and M. M. Sani, "Internet Traffic Analysis With Goodness Of Fit Test On Campus Network," ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 21, pp. 9892–9898, 2015.
- [16] H. El-sallabi, M. Abdallah, and K. Qaraqe, "Modelling of Parameters of Rician Fading Distribution as a Function of Polarization Parameter in Reconfigurable Antenna," 2014 IEEE/CIC International Conference on Communications in China (ICCC), pp. 534–538, 2014.