

Packet Inter-arrival Time Distribution in Academic Computer Network

E. Garsva¹, N. Paulauskas¹, G. Grazulevicius¹, L. Gulbinovic¹

¹*Department of Computer Engineering, Vilnius Gediminas Technical University, Naugarduko St. 41–440, LT-03227 Vilnius, Lithuania*

Abstract—The article presents the statistical analysis results of network packet inter-arrival time distribution in academic computer network. Most popular transport protocols TCP and UDP are addressed in the research. Data was gathered using NetFlow protocol. Network traffic was divided into sections according its direction and usage trends, then packet inter-arrival time distributions were found. Kolmogorov-Smirnov test was used to evaluate goodness-of-fit of packet inter-arrival time distributions and it was determined, that Pareto Second Kind distribution fits the majority of the experimental distributions.

Index Terms—Computer networks, packet inter-arrival time distribution, statistical analysis, statistical distribution.

I. INTRODUCTION

Computer networks modelling or performance evaluation requires knowledge of the computer network characteristic distribution trends. Distribution according to the known statistical law is very applicable during the modelling and simulation, but the uncertainty remains: does the statistical distribution represent the real situation in the production computer network. This research addresses computer network of Electronics faculty of Vilnius Gediminas Technical University (VGTU). The source for getting the information about the network usage is the monitoring system which uses NetFlow protocol. NetFlow protocol originally developed by Cisco Systems Incorporated became an industry standard for IP traffic monitoring.

The article presents the research of TCP and UDP packet inter-arrival time distributions. The difficulty here is that packet's inter-arrival time is not directly available from the NetFlow data. Only the time interval between separate flows is available, so we made an assumption that all the packets in any NetFlow are distributed with the uniform time intervals and then used this data to find the distribution which best fits the experimental data. This is a first and essential step in network traffic modeling and simulation. We found that Pareto Second Kind distribution fits best in most of the cases, so it can be used to model packet inter-arrival time.

II. RELATED WORK

The study of computer network packet inter-arrival time distribution is very important for network traffic simulation and modeling. The accurate modeling and simulation of such traffic is a fundamental part of packet switching network

design, resource allocation and management [1].

Packet inter-arrival time distribution was described as Poisson distribution at first, but problems have been identified over time with this model [2]. Network traffic has shown that it is often bursty over a wide range of time scales.

Network traffic model for non-congested Internet backbone links based on the collected network flows can be represented by Poisson short-noise process and is able to find a good approximation for the average traffic on backbone link and its variations at short time scale [3]. Monitoring of network edges is enough to dimension the backbone links. The model can help to optimize the utilization of the available resources in the backbone and to evaluate the impact caused by a change in flow durations or due to the number of user increase in the congested access networks.

Local Area Network traffic like Wide Area Network due to its heterogeneity can be modeled by heavy tail distributions [4]. Various parameters of the captured network traffic are investigated and characterized; this led to conclusion that packet inter-arrival time follows power law and can be modeled by Pareto distribution.

Lognormal distribution models network traffic inter-arrival time better than Pareto distribution, especially, in the lower tail region [5]. That was determined by statistical analysis of the network traffic inter-arrival time. Primary reason attributed for this observation is, exponential back-off approach in the Ethernet layer and greediness inherent into the protocol leading to the competition amongst senders and resulting into the capture effect. The authors gathered the packets from the network and performed further analysis using the packet data. The research presented in our article uses the data acquired from NetFlow.

There are a number of widely used network traffic models, with different features and traffic characteristics they capture best. This leads to a conclusion that, there is no single model that can be used effectively to model the traffic in all kinds of networks. Standard goodness-of-fit tests such as Kolmogorov-Smirnov, Anderson-Darling, Chi-Square or modifications of theirs [6] allow an optimal and mathematically proven network traffic distribution to be gotten. The easiest traffic generation model composition would be the single distribution with the tunable coefficients.

III. NETWORK TRAFFIC STATISTICAL ANALYSIS

Computer network used in this research is the same as in

our previous one [7] and the main differences are: NetFlow data was acquired on the October 2012 and only working days are considered, because the number of network packets during weekends is small and makes up only 7.3 % of all packets during a month.

Statistical packet data was extracted from the NetFlow data assuming that all packets in any network flow are distributed with the uniform time intervals. Packets are transmitted one after another, but several network flows can occur at the same time. After sorting the arrival times of packets from each flow, different and randomly distributed packet inter-arrival times are obtained.

This research is focused on the packet inter-arrival time distribution, so packet size, source and destination ports as well as IP addresses are not taken into account.

Average daily network traffic, calculated from the average values of the exact period of the day during the month for both incoming and outgoing traffic, represented in the number of flows N_F is presented in stacked area graph Fig. 1.

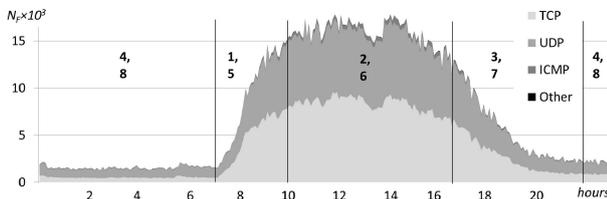


Fig. 1. Average daily network traffic in number of flows: TCP, UDP, ICMP and Other.

Network traffic graph is divided into 4 sections according to the network usage: network is not used by the users during the night, only some scheduled tasks are performed during the time 22.00–7.00 (4, 8), network is used as intended during the day 10.00–16.30 (2, 7) and network usage start to rise 7.00–10.00 (1, 5) and falls 16.30–22.00 (3, 7) when users come and go into and from the faculty. Network traffic can be described according to the direction: incoming traffic (1–4) is the one which comes into the network from outside networks and outgoing (5–8) is the one which originates from the network. Network traffic sections are described in Table I.

TABLE I. NETWORK TRAFFIC SECTIONS.

Incoming	Packets, %	Time	Length, h	Packets, %	Outgoing
1	6.96	7.00–10.00	3	4.36	5
2	36.58	10.00–16.30	6.30	23.05	6
3	16.48	16.30–22.00	5.30	10.78	7
4	0.79	22.00–7.00	9	1.01	8

Network traffic sections are different in length. The peak time (2, 6) is the second longest but contains the biggest part of packets and night time (4, 8) is the longest, but the least amount of packets is monitored. Network usage rises (1, 5) a bit faster than it falls (3, 7) but there are 2.4 times more packets in the falling traffic section than in the rising one. Network traffic statistics according to transport protocols is presented in Table II and Table III for incoming and outgoing traffic accordingly. Group Other contains various routing and tunneling protocols which do not use first three protocols for transport.

Data presented in Tables II and III does not take into account 6.4 % of packets which fall into the edge of traffic

sections, due to the fact that nfdump tool does not process the flows which are started but not ended during the set time window.

TABLE II. INCOMING NETWORK TRAFFIC STATISTICS.

Parameter \ Protocol	TCP	UDP	ICMP	Other
Traffic total, B $\times 10^9$	843	257	0.158	36
Flows total	12954137	11733655	438532	87850
Packets total, $\times 10^6$	1007	265	2	30
Average packets in flow	78	23	4	339
Average size of packet, B	838	970	96	1197

TABLE III. OUTGOING NETWORK TRAFFIC STATISTICS.

Parameter \ Protocol	TCP	UDP	ICMP	Other
Traffic total, B $\times 10^9$	131	91	0.358	3
Flows total	12072179	10958168	882523	35742
Packets total, $\times 10^6$	464	169	3	14
Average packets in flow	38	15	3	384
Average size of packet, B	282	539	125	228

Incoming traffic dominates as there is 6.4 times more TCP and 2.8 times more UDP incoming data, number of flows is only 7 % higher for both protocols and the number of packets is 2.2 times more for TCP and 1.6 for UDP. Average number of packets in a flow is higher for incoming traffic, for both protocols, but the difference is set by the packet size which is 3 times bigger for TCP and 1.8 times for UDP.

There is 2.8 times more TCP data than UDP, number of flows is only 1 % higher for TCP and the average number of packets in a flow is 3 times higher for TCP. UDP average packet size is 26 % bigger than TCP.

The amount of flows, packets and data of ICMP and Other transport protocols is significantly lower than TCP and UDP and is not addressed in this research.

Academic computer network is based on Ethernet protocol, with 100 Mbps and 1 Gbps segments. 100 Mbps Ethernet has 0.96 μ s minimum inter frame gap and 1 Gbps Ethernet minimum inter frame gap equals 0.096 μ s. The packet size in Ethernet network can be between 64 and 1518 Bytes. Because of this in 1 Gbps network segment minimal inter-arrival time between packets varies because of the packet size between 0.608 μ s and 12.240 μ s. For 100 Mbps network minimal packet inter-arrival time will be from 6.080 μ s to 122.400 μ s. The granularity used in this research 0.1 ms was chosen knowing that the average inter-arrival time in the network is 2.835 ms. It was determined that granularity of 1 ms to be not enough as ~80 % of values felt into the first interval. Granularity higher than 0.1 ms is complicated as NetFlow uses 1 ms granularity, and 0.1 ms granularity is possible because one NetFlow usually consists of more than one packet.

IV. NETWORK PACKET INTER-ARRIVAL TIME DISTRIBUTION

The network of the faculty is not overloaded, so it is expected, that the biggest amount of traffic arrives when it is needed the most, incoming traffic is more intense. Outliers were introduced to eliminate a situation when unitary peaks of network traffic influence the whole network load.

Without the outliers the situation might rise when during some set time interval the number of packets is 10^3 , but during one of the days the number of packets during the same time interval reaches 10^5 , that single day would impact the final result. Quartiles Q_1 and Q_3 are found, interquartile

range IQR is calculated and then all the values falling into the interval between the lower fence and upper fence are considered (1)

$$[Q_1 - 3 \cdot IQR, Q_3 + 3 \cdot IQR], IQR = Q_3 - Q_1. \quad (1)$$

Network packet inter-arrival time was split into intervals: $((n-1) \times 10^{-4}, n \times 10^{-4}), n \in [1, 5000]$.

Higher number of the intervals is not used due to the low number of the packets falling into them. Presented graphs plot the curves to the amount it reveals the trend, cutting the long tails. Total number of inter-arrival time entries for one transport protocol is 23 (working days) $\times 5000$ (number of intervals) = 115000 (entries). Outlier is a particular interval for a particular day which does not fall into the fenced interval (1) and is discarded. The percentage of the outliers and the percentage of packets which belong for both transport protocols are presented in Table IV.

TABLE IV. OUTLIER PERCENTAGE.

	TCP		UDP	
	Outliers, %	Packets, %	Outliers, %	Packets, %
1	3.80	5.07	4.22	56.09
2	1.67	2.26	5.20	52.04
3	4.58	7.56	5.37	21.58
4	3.61	47.91	3.62	15.92
5	4.17	6.95	2.13	36.66
6	1.43	4.44	2.27	11.53
7	4.40	12.16	2.63	34.84
8	3.47	56.12	4.71	30.98
Incoming	3.41	15.70	4.60	36.41
Outgoing	3.37	19.92	2.94	28.50
Average	3.39	17.81	3.77	32.46

3.39 % of outliers were found for TCP and 3.77 % for UDP that results into 17.81 % of TCP and 32.46 % of UDP packets being not taken into account. The number for outgoing traffic outliers is lower for TCP – 0.04 % (despite, the fact that number of packets is 4.22 % higher) and for UDP – 1.66 %.

The highest number of outliers for TCP is in traffic section 3 %– 4.58 % (7.56 % of packets) and for UDP it is traffic section 8 %– 4.71 % (30.98 % of packets). Packet wise situation differs: for TCP it is traffic section 8 (56.12 % of packets fall into 3.47 % of outliers) and for UDP it is traffic section 1 (56.09 % packets fall into 4.22 % outliers).

Computer network with low load introduces big number of outliers because of the pulses, while heavily loaded network tend to have less outliers. UDP introduces more outliers than TCP.

Distribution of time interval $f(n)$ between packets for different traffic sections is presented in Fig. 2 and Fig. 3. n represents the end of interval between the packets.

The highest values are for traffic sections 2 and 6 which represent the peak of network traffic and lowest values are for traffic sections 4, 8 which represent night period, this is true for TCP traffic and partially true for UDP, as traffic section 3 outstands.

Packet inter-arrival time distributions directly depend on the number of packets and Cumulative Distribution Function (CDF) $F(n)$ normalizes the trends, so the number of packets does not overwhelm the trend. Graphs for TCP and UDP are presented in Fig. 4 and Fig. 5 respectively.

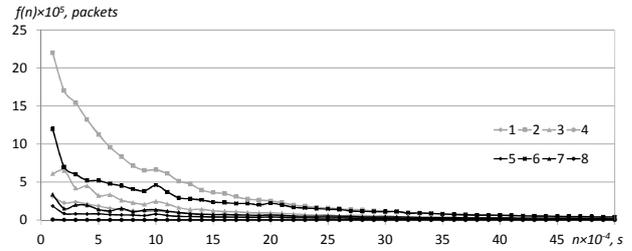


Fig. 2. Distribution of TCP packet inter-arrival time.

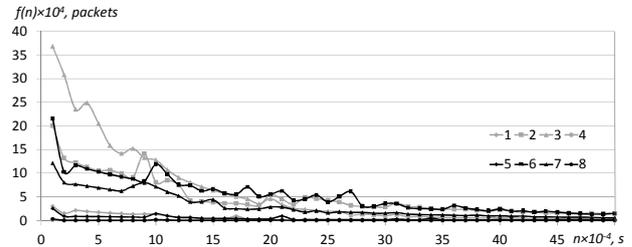


Fig. 3. Distribution of UDP packet inter-arrival time.

CDF of TCP packet inter-arrival time shows that 70 % of all the TCP packets in network sections 1, 3 and 6 arrive in less than 2 ms and 80 % of packets arrive during the same time interval when the usage is the highest for network section 2. 50 % of UDP packets in network sections 1, 2 and 6; 60 % in section 7; 80 % in section 3 arrive in less than 2 ms (Fig. 5).

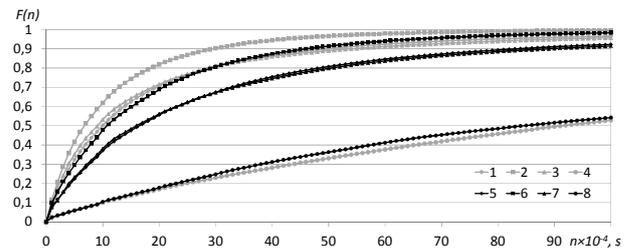


Fig. 4. CDF of TCP packet inter-arrival time.

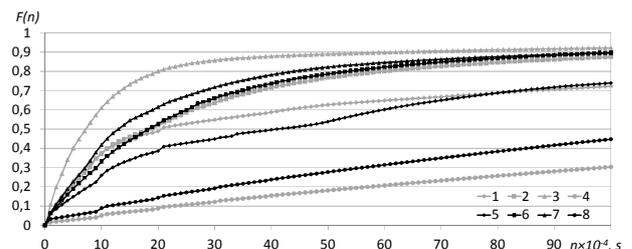


Fig. 5. CDF of UDP packet inter-arrival time.

Best fitting distributions were determined using Kolmogorov-Smirnov goodness-of-fit test. Weibull, Pareto, Gamma, Exponential and Lognormal distributions were considered as they are close to an average experimental distribution curve and are used in computer network traffic modelling. The goodness-of-fit tests were performed for packet inter-arrival time distributions based on the average network traffic going both directions (Table V). Weibull, Pareto, Lognormal and Gamma distributions use shape parameter and scale parameter. Exponential distribution uses rate parameter. A shape parameter affects the shape of a distribution and scale parameter stretches or shrinks it. The first parameter of distribution in Table V is represented by Param. 1 and the second by Param. 2. Kolmogorov-Smirnov

parameter KS shows the maximum absolute difference between the experimental and distribution curves and the lower it is the better is the fit. Multiplier A is used to adjust the fitted distribution to the experimental curve on y axis and it changes the value of Probability Density Function integral to the value of A .

TABLE V. GOODNESS-OF-FIT OF PACKET INTER-ARRIVAL TIME DISTRIBUTION.

Protocol	Distribution	Param. 1	Param. 2	KS	A	Rank
TCP	Weibull	0.798	0.002	0.063	609.64	2
	Pareto 2	2.658	0.003	0.058	606.03	1
	Gamma	0.509	0.004	0.129	638.19	-
	Exponential	561.680	-	0.123	603.03	-
	Lognormal	1.409	-7.151	0.075	602.14	3
UDP	Weibull	0.734	0.003	0.053	322.35	-
	Pareto 2	1.848	0.004	0.046	322.90	1
	Gamma	0.480	0.008	0.103	333.79	-
	Exponential	249.330	-	0.176	317.91	-
	Lognormal	1.545	-6.472	0.053	323.57	2

For the average network traffic Pareto Second Kind (Pareto 2) distribution fits best for both protocols as it is seen in Fig. 6.

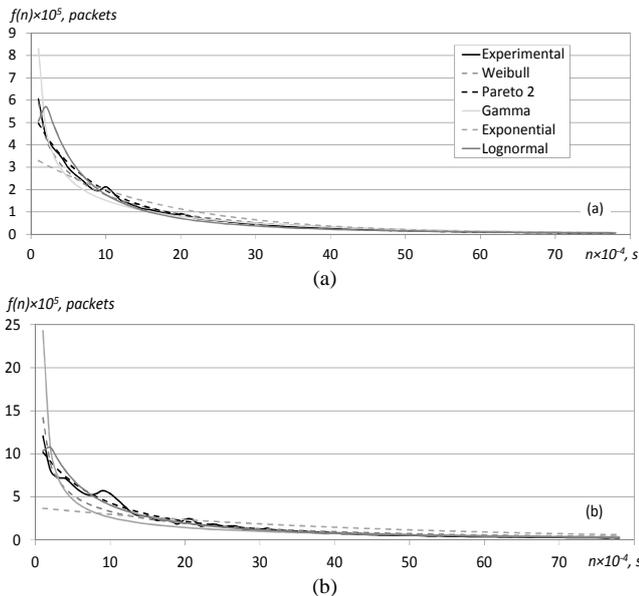


Fig. 6. Goodness-of-fit of packet inter-arrival time distributions: (a) TCP, (b) UDP.

Pareto Second Kind distribution is a standard Pareto distribution with shifted x axis so it falls into $0 < x < +\infty$, while in standard Pareto distribution $x < +\infty$. Pareto Second Kind probability distribution conditionally can be called Lomax distribution and is a heavy-tail distribution usually used in business or economical modelling.

Network traffic sections are considered and goodness-of-fit was performed for all the sections in order to determine its distributions and parameters. CDFs are considered up to 0.99, as other values contribute only to the tails.

All distribution curves for TCP were successfully fitted; fitting UDP curves was challenging (Table VI): 1 and 5 curves were not fitted; best values were presented in the table.

Pareto 2 fits most of the curves, especially those where the traffic is growing or falling. Pareto 2 distribution was the best or second best for all the distributions except TCP

section 6, so Pareto 2 distribution can be chosen to model network packet inter-arrival time.

TABLE VI. PACKET INTER-ARRIVAL TIME DISTRIBUTION COEFFICIENTS.

Protocol	Section	Distribution		KS	A	
TCP	1	Pareto2	2.7278	0.0034	0.0478	359.63
	2	Weibull	0.8978	0.0011	0.0623	1890.70
	3	Pareto2	1.9981	0.0022	0.0558	698.54
	4	Pareto2	3.1931	0.0361	0.0246	23.18
	5	Weibull	0.7750	0.0027	0.0441	230.12
	6	Gamma	0.7810	0.0023	0.0545	1201.46
	7	Weibull	0.7578	0.0028	0.0467	437.86
	8	Weibull	0.7897	0.014	0.0211	24.93
UDP	1	Lognormal	1.9497	-6.084	0.0560	46.55
	2	Pareto2	1.8484	0.0039	0.0456	322.91
	3	Pareto2	1.2351	0.0009	0.0643	339.50
	4	Pareto2	4.0406	0.1065	0.0221	10.35
	5	Weibull	0.6066	0.0073	0.0476	40.96
	6	Pareto2	1.9606	0.0042	0.0409	338.45
	7	Pareto2	1.4612	0.0022	0.0454	184.25
	8	Pareto2	2.7942	0.0386	0.0293	10.15

V. CONCLUSIONS

- Network traffic division into sections is reasonable, as it reveals dominant trends which are needed in order to compose general network model and to choose definitive statistical distribution.

- Outliers conclude 3.6 % for TCP as well as UDP, but the number of packets which fall into those outliers depends on day time and network utilization: most of such packets are during the night time when network utilization is low for TCP, for UDP incoming traffic has more of such packets and the lower network utilization lowers its number.

- Kolmogorov-Smirnov goodness-of-fit test shows that Pareto Second Kind distribution fits best for both TCP and UDP network packet inter-arrival time distribution experimental curves, this also show that TCP and UDP network packet inter-arrival time distributions are of the same shape.

REFERENCES

- [1] V. Grout, S. Cunningham, D. Oram, R. Hebblewhite, "A Note on the Distribution of Packet Arrivals in High-Speed Data Networks", in *Proc. IADIS Int. Conf. WWW/Internet 2004*, Oct. 6–9, 2004, pp. 889–892.
- [2] V. Paxson, S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–244, June 1995. [Online]. Available: <http://dx.doi.org/10.1109/90.392383>
- [3] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, P. Owezarski, "Modeling internet backbone traffic at the flow level," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2111–2124, Aug. 2003. [Online]. Available: <http://dx.doi.org/10.1109/TSP.2003.814521>
- [4] S. A. Mushtaq, A. Rizvi, "Statistical analysis and mathematical modeling of network (segment) traffic", in *Proc. IEEE Symposium on Emerging Technologies*, 2005, pp. 246–251.
- [5] A. Bhattacharjee, S. Nandi, "Statistical analysis of network traffic inter-arrival", in *Proc. IEEE 12th Int. Conf. Advanced Communication Technology (ICACT)*, Feb. 7–10, 2010, vol. 2, pp. 1052–1057.
- [6] M. Fras, J. Mohorko, Z. Cucej, "A new goodness of fit test for histograms regarding network traffic packet size process", in *Proc. IEEE Int. Conf. Advanced Technologies for Communications (ATC)*, 2008, pp. 345–348.
- [7] E. Garsva, N. Paulauskas, G. Grazulevicius, L. Gulbinovic, "Academic Computer Network Traffic Statistical Analysis", in *Proc. of the IEEE 2nd Baltic Congress on Future Internet Communications (BCFIC)*, 2012, pp. 100–105.