VOIP TRAFFIC SHAPING ANALYSES IN METROPOLITAN AREA NETWORKS

Rossitza Goleva, Mariya Goleva, Dimitar Atamian, Tashko Nikolov, Kostadin Golev

Abstract: This paper represents VoIP shaping analyses in devices that apply the three Quality of Service techniques – IntServ, DiffServ and RSVP. The results show queue management and packet stream shaping based on simulation of the three mostly demanded services – VoIP, LAN emulation and transaction exchange. Special attention is paid to the VoIP as the most demanding service for real time communication.

Keywords: Packet network, IP, Quality of Service, VoIP, shaping.

ACM Classification Keywords: C.4 Performance of Systems, C. Computer Systems Organization, C.2 Computer-Communication Networks

Introduction

IP networks and their Quality of the Service are challenging area for investigation. In spite of the fact that they are easy for use, enough cheap and quite useful in human life there is recently high demand for IP network use instead of all other kind of communication. Real time and non real time services and applications interwork on the same infrastructure. Different services have different quality requirements. The quality offered by the network depends on the traffic. In this paper we analyze the traffic shaping effect of the three mostly used techniques – IntServ, DiffServ, and RSVP. The analyses are made on the basis of the three popular services – VoIP, LAN emulation, transaction exchange [Jha], [Janevski], [Pitts], [Ralsanen]. The shaping effect is estimated under typical queueing circumstances. The model uses queues and priorities specific for the IntServ, DiffSerf, and RSVP. The reason is to investigate the effect that can be reached without implementation of the expensive shaping devices. This fractional shaping phenomenon is important in small to medium wire and wireless Metropolitan Area Networks (MAN) that grow rapidly. Changing circumstances in ad hoc networks also can apply the results presented.

Traffic sources

The traffic sources generate combination of three types of services in the network – Voice over IP, LAN emulation and transaction exchange. The size of the example network is typical for the business area. Some assumptions are made for every traffic source. In Voice over IP (VoIP) service silence and talk intervals are exponentially distributed with equal mean values [Jha], [Pitts]. There are authors who use talk to silence ratio of ½. Others do prefer to use on-off model for voice service. The behavior of the VoIP end-user is supposed to be similar to the phone user. The limits for waiting times are calculated under consideration of end-to-end delay bounds for every service [Lavenberg], [Iversen]. The same is valid for queue length. Servicing times per packets are fixed for LAN connection of 100 Mbps. Table 1 represents traffic sources parameters in the model.

LAN emulation is modeled with sessions. Sessions are established for any Internet connections. Packet rate is higher in comparison to the VoIP. Session duration is low. The traffic source is behaving as on-off model with exponential duration of the silence and transmission intervals [Lavenberg]. Transaction exchange is specific with few packets exchange. The service is not time demanding. Sessions are short and similar to the datagram exchange.

Table 1. Traffic Sources Parameters

No	Parameter	VoIP	LAN emulation	Transactions
1.	Pear rate, packets per second	10	164	0
2.	Mean call/ session duration, sec	180	20	10
3.	Mean duration between calls/sessions, sec	360	10	15
4.	Mean talk/ silence duration, sec	20	5	2
5.	Distribution of call/series duration	Exponential	Exponential	Exponential
6.	Maximal waiting time, sec	0.00072	0.6	1

7.	Maximal number of waiting packets	210	1804	2
8.	Traffic sources	5000	500	1500
9.	Priorities	High	Medium	Low
10.	Packet length, bytes	800	800	800

Number of traffic sources is taken from the typical business area. Packets are taken to be long. IP packets of 800 bytes carry up to 80 milliseconds voice. This means that quality voice can be transmitted only in the area with up to 2-3 hops. Therefore, we design VoIP service for regional connectivity. More precision investigation can be done with up to 200 bytes voice packets.

Integrated Services

Integrated Services (IntServ) is a complex technique that ensures Quality of Service in IP networks. It is applied usually in access routers or gateways and tried to serve packets from different services in a different ways depending on the quality requirements. IntServ classifies services into three main classes depending on the traffic requirements [Janevski]:

- Elastic application;
- Tolerant real-time applications;
- Intolerant real-time applications.

Elastic applications are served with “best effort” discipline [Tanenbaum]. They are served without any guarantee of quality level like transaction exchange. Tolerant real-time applications are delay sensitive and usually require high bandwidth. Token bucket model with peak rate control is a proper model for such traffic. LAN emulation is usually modeled this way. Some authors propose token bucket that controls series length and mean rate for more accuracy. Many authors propose the two token buckets to be connected in a cascade as it is shown on Figure 1 [Ralsanen]. Intolerant real-time applications require low delays and almost guaranteed bandwidth. The model with two cascaded token buckets is compulsory for such traffic [Jha]. VoIP service is intolerant to the quality degradation service. IntServ simulation model is based on two cascaded token buckets that bound peak rate, series length and mean rate of the traffic (Figure 1). The model is approximated as a black box that changes the characteristics of the data at output in specific for IntServ way. As a result after approximation and few calculations it is easy to derive simpler model with one FIFO queue, priorities, fixed rate at the output and different limits for waiting times in the queue. The resulting model is represented on Figure 2. This is the model that has been simulated further. Table 2 represents main data for model behavior.

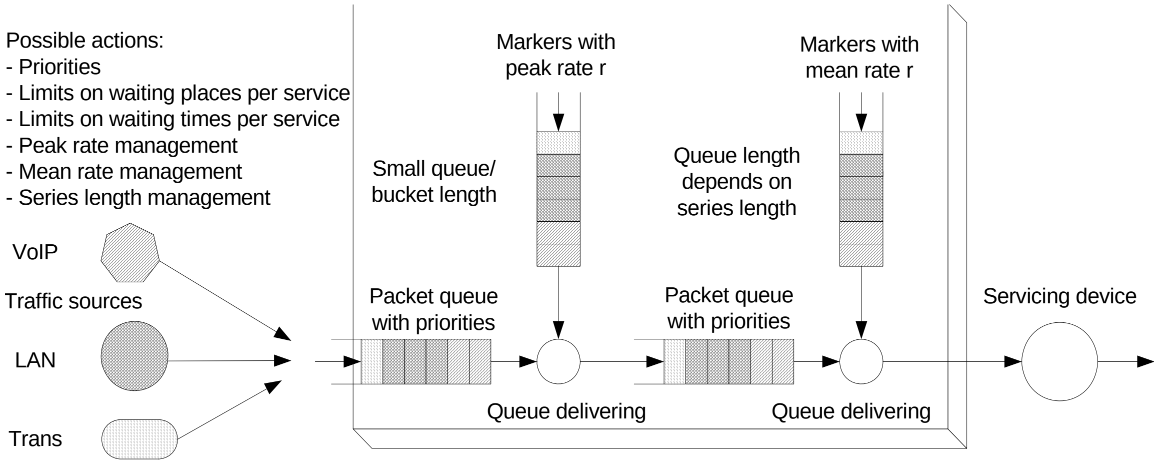

Figure 1. Black box IntServ model approximation

Differentiated Services

Differentiated Services (DiffServ) is another quality management technique that is more applicable for core networks. Due to its nature DiffServ applies its rules on aggregated traffic. After appropriate marking of the aggregated packets they are gathered in the way that is defined for their class. There are three main types of services we highlighted in this paper [Pitts]:

- Premium service with low delay, low loss, guaranteed bandwidth like VoIP;
- Assured service with less requirements to the delay and loss in comparison to the premium service like LAN emulation;
- Olympic service with no time requirements at all like transaction exchange.

The model from Figure 2 with different parameters is used to represents DiffServ application. The parameters are shown on Table 2.

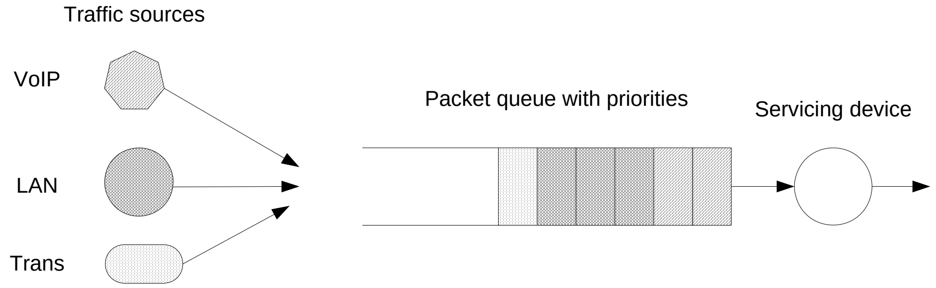

Figure 2. Final IntServ model with input data, bounds for waiting times and queue length specific for service type

RSVP

Resource Reservation Protocol (RSVP) is a technique useful for delay sensitive traffic like VoIP. Three types of services are identified for RSVP like:

- Wildcard filter that is applied to gather maximal requirements for given interface like LAN emulation;
- Shared explicit that is applied to gather maximal requirements for the interface taking into account called address. Transaction exchange is modeled as shared explicit service;
- Fixed filter that requires full reservation for quality sensitive services like VoIP.

The model simplified for IntServ and DiffServ procedures is applied with specific parameters for RSVP. Characteristics of the derived model are shown on Table 2.

Table 2. Model characteristics

No	Parameter	IntServ	DiffServ	RSVP
1.	Queue length, packets	2016	1840	1840
2.	VoIP queue length fraction, packets	210	200	200
3.	LAN queue length fraction, packets	1804	1640	1640
4.	Transaction queue length fraction, packets	2	2	2
5.	Maximal waiting time for VoIP, sec	0,000716	0,0303	0,07508
6.	Maximal waiting time for LAN, sec	0,6	0,27876	0,69
7.	Maximal waiting time for transactions, sec	1	1	1
8.	Priority for VoIP	Highest	Highest	Highest
9.	Priority for LAN	Medium	Medium	Medium
10.	Priority for transactions	Low	Low	Low

Results

Simulation is performed on C++ language. The pseudo exponential pseudo deterministic characteristics of the traffic sources are reached after usage of combination between many random generators [Kleinrock], [Iversen], [Lavenberg]. The queue behavior is complex due to the priorities and limits for waiting times. Many parameters have been derived from the model like time and space loss probabilities, probabilities to wait for different types of traffic, statistical data for probability distribution functions and probability density functions of the packets intervals, queue lengths, waiting times at many interface points of the model like output of the traffic sources, input and output of the queue. Statistical accuracy of the derived results is proven by Student criterion. IntServ, DiffServ and RSVP have different way to gather with packets and this influences the way they drop packets and shape them.

On Figure 3, 4 and 5 observations of packet intervals at the input and output of the queue are shown. It is interesting for shaping estimation. The effect of fast servicing in RSVP can be seen from Figure 3. The delay variation of the packet intervals is becoming smoother and tends to constant value. Similar result is visible for IntServ on Figure 4. On Figure 5 shaping of the IntServ and DiffServ is seen.

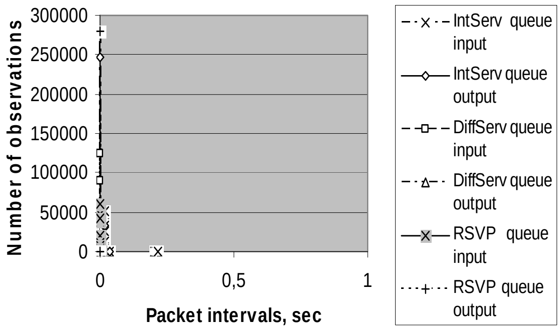

Figure 3. Delay variation reduction in RSVP

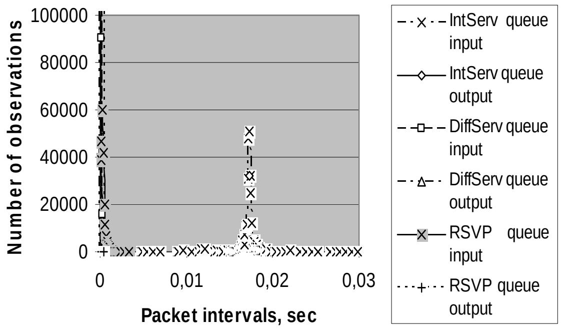

Figure 4. Delay variation reduction in IntServ

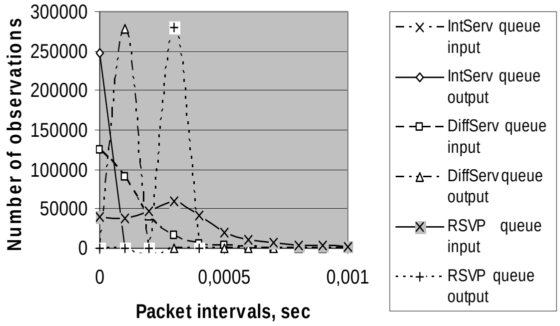

Figure 5. Delay variation reduction in IntServ and DiffServ

Interesting results that influence directly interfaces and queue management are derived on the basis of queue length per service type. The queue fraction of the three services is observed. It is visible from Figure 6 that for services with highest priority like VoIP IntServ is the most proper shaping mechanism. With some not quite accurate approximation the distribution of the queue length can be considered exponential. Figure 7 represents the observations for LAN service. Because of the less critical waiting times and low priority the distribution tends to be deterministic.

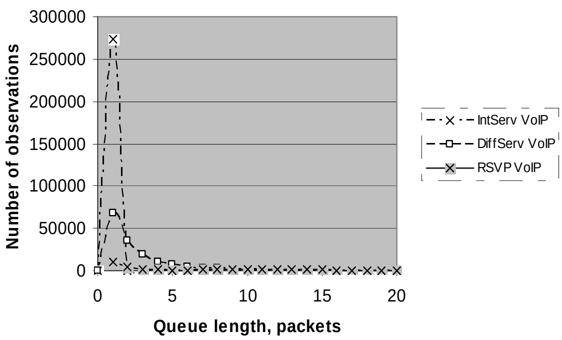

Figure 6. Observations of queue length in VoIP

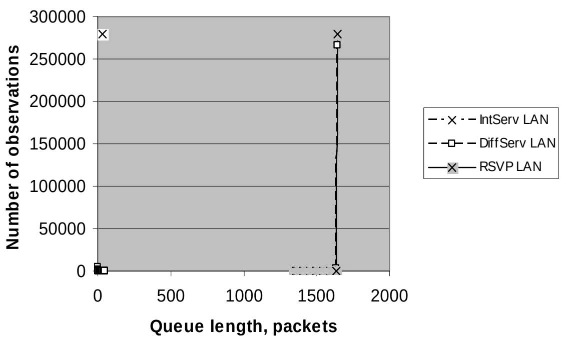

Figure 7. Observations of queue length in LAN

On Fugure 8 and 9 the observations of packet intervals only between voice packets are shown. The statistical multiplexing effect and shaping phenomenon are due to the high priority of the voice traffic in comparison to the priority of the data traffic. On Figure 10 the shaping effect of the three techniques is visible. On Figure 11 and 12 only effect of DiffServ is obvious in different observation scales.

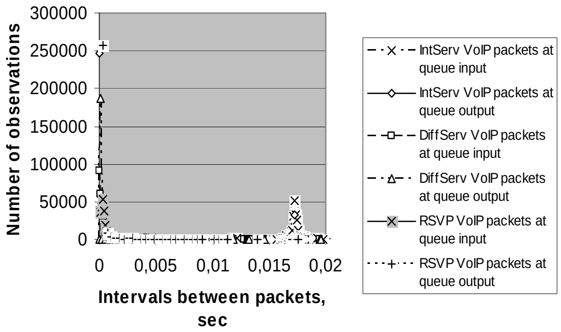

Figure 8. Observations of intervals between VoIP packets

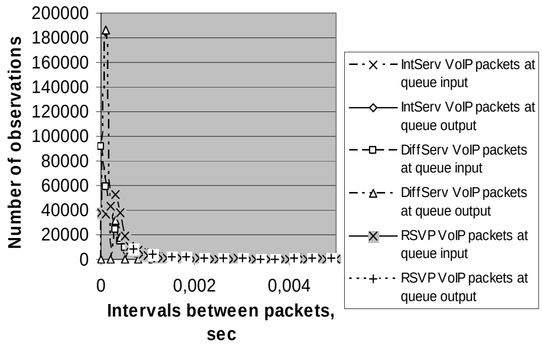

Figure 9. Observations of intervals between VoIP packets for DiffServ

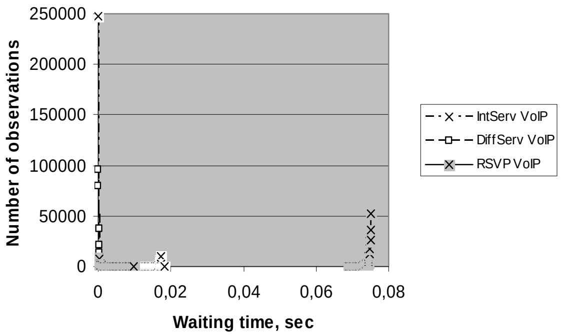

Figure 10. Observations of waiting times for VoIP packets

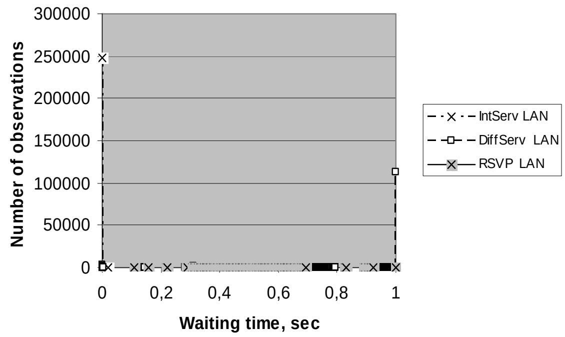

Figure 11. Observations of waiting times for LAN packets

Conclusion

In this paper we show observations of the packet intervals at the queue input and queue output as well as statistical data of queue length, waiting times and loss per service type (Table 3). These results demonstrate the specific characteristics of the queue as a packet shaper in three QoS management algorithms IntServ, DiffServ, RSVP. The shaping effect is possible for priority service types.

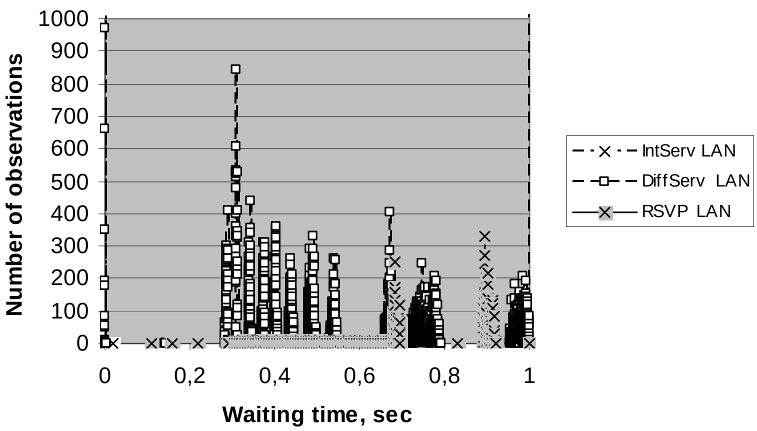

Figure 12. Observations of waiting times for LAN packets

The low delays for priority services types are due to the bigger delays for non priority service types. The waiting times are redistributed due to the QoS algorithm and priority. The deterministic nature of the packet streams suppress shaping and increase losses. The statistical multiplexing effect is very limited due to the deterministic streams. Mean values of the queue lengths, probability to wait, loss probability due to the lack of space and waiting time bounds per discipline and per service redistribution are visible from Table 3. They can be used for configuration planning of the time and space limits in the router interfaces.

The results demonstrate the capability of IntServ to define excellent service for its higher priority applications. It is promising in access networks. DiffServ shows excellent resource management and utilization and therefore is better for core services. RSVP is a good counterpart of IntServ in access networks.

The authors refine the simulation model with more traffic sources and more precise generation of the packets from these sources based on the observation of the real traffic. MMPP and geometric/ Weibull distributions are also considered. Limits criteria for queue management are under investigation.

Table 3. Queue length, waiting times, loss probability

Mechanism	IntServ	DiffServ	RSVP
Overall mean queue length, packets	37.97523	1586.961	1798.409
Mean queue length of VoIP fraction, packets	1	2.24207	156.9017
Mean queue length of LAN fraction, packets	35	1583.561	1639.675
Mean queue length of Trans fraction, packets	2	1.98331	2
Overall loss probability due to the lack of space	0.0094	0.77815	0.91222
VoIP packets loss probability due to the lack of space	0	0	0.33540
LAN packets loss probability due to the lack of space	0	0.89373	0.98747
Transaction packets loss probability due to the lack of space	1	0.99574	1
Overall loss probability due to waiting time bound	0.98865	0	0
VoIP packets loss probability due to waiting time bound	0.98109	0	0
LAN packets loss probability due to waiting time bound	1	0	0
Transaction packets loss probability due to waiting time bound	0	0	0
Overall probability to wait	0.00191	0.22074	0.08767
VoIP packet probability to wait	0.0189	0.9996	0.66433
LAN packet probability to wait	0	0.105130	0.01244
Transaction packet probability to wait	0	0.00339	0
Overall interface occupancy, fraction	0.13668	0.13644	0.13704
Interface occupancy due to the VoIP traffic, fraction	0.13621	0.05046	0.08955
Interface occupancy due to the LAN traffic, fraction	0.00048	0.08588	0.04749
Interface occupancy due to the transaction traffic, fraction	0	0.00009	0

Acknowledgements

This paper is sponsored by the Ministry of Education and Research of the Republic of Bulgaria in the framework of project No 105 “Multimedia Packet Switching Networks Planning with Quality of Service and Traffic Management”.

Bibliography

[Jha, 2002] Jha, S., M. Hassan, “Engineering Internet QoS”, Artech Hose, 2002.

[Janevski, 2003] Janevski, T., “Traffic Analysis and Design of Wireless IP Networks”, Artech House, 2003.

[Kleinrock, 1976] Kleinrock, Leonard, “Queueing Systems”, Volumes I and II, John Wiley and Sons, 1976.

[Iversen, 2005] Iversen, V., “Teletraffic Engineering Handbook”, ITU-D, 2005.

[Lavenberg, 1983] Lavenberg, Stephen S., Editor, “Computer Performance Modeling Handbook”, Academic Press, 1983, ISBN 0-12-438720-9.

[Pitts, 2000] Pitts, J., J. Schormans, “Introduction to IP and ATM Design and Performnce”, John Wiley&Sons, Ltd., 2000.

[Ralsanen, 2003] Ralsanen, V., “Implementing Service Quality in IP Networks”, John Wiley & Sons, Ltd., 2003.

[Tanenbaum. 2003] Tanenbaum, Andrew S., “Computer Networks, Second Edition”, Prentice-Hall International, Inc., 2003, ISBN 0-13-166836-6.

Authors' Information

Rossitza Iv. Goleva – Assistant-Professor; Department of Telecommunications, Technical University of Sofia, Bulgaria, Kl. Ohridski blvd. 8, Sofia. 1756, Bulgaria; e-mail: rig@tu-sofia.bg

Mariya At. Goleva – Student; Department of Communication Networks, University of Bremen, Germany; e-mail: mgoleva@gmail.com

Dimitar K. Atamian - Assistant-Professor; Department of Telecommunications, Technical University of Sofia, Bulgaria, Kl. Ohridski blvd. 8, Sofia. 1756, Bulgaria; e-mail: dka@tu-sofia.bg

Tashko Nikolov – Assistant-Professor, Ph.D. Department of Telecommunications, Technical University of Sofia, Bulgaria, Kl. Ohridski blvd. 8, Sofia. 1756, Bulgaria; e-mail: tan@tu-sofia.bg

Kostadin At. Golev – Developer; Bianor Ltd., Bulgaria; e-mail: kotseto@gmail.com

STUDY OF QUEUEING BEHAVIOUR IN IP BUFFERS

Seferin Mirtchev

Abstract: It is unquestioned that the importance of IP network will further increase and that it will serve as a platform for more and more services, requiring different types and degrees of service quality. Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. In this paper, we investigate the queueing behaviour found in IP output buffers. This queueing increases because multiple streams of packets with different length are being multiplexed together. We develop balance equations for the state of the system, from which we derive packet loss and delay results. To analyze these types of behaviour, we study the discrete-time version of the "classical" queue model M/M/1/k called Geo/Gx/1/k, where Gx denotes a different packet length distribution defined on a range between a minimum and maximum value.

Keywords: delay system, queueing analyses, discrete time queue, IP traffic modelling; packet size distribution.

ACM Classification Keywords: G.3 Probability and statistics: queueing theory, I.6.5 Model development

Introduction

The initial motivation for this paper is the necessity of traffic engineering in IP networks. Many analyses of Internet traffic behaviour require accurate knowledge of the traffic characteristics for purposes ranging from a management of the network quality of service to modelling the effect of new protocols on the existing traffic mix.

Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. The proper functioning of these protocols requires an increasingly detailed knowledge for statistical characteristics of IP packets. The amount of information flowing through the network also increases, and the challenge is to obtain the accurate information from a huge set of data packets.

The packet queueing in an IP router arises because multiple streams of packets from different input ports are being multiplexed together over the same output port. A key characteristic is that the packets have different length. The minimum header size in IPv4 is 20 octets, and in IPv6, it is 40 octets. The maximum packet size depends on the specific sub-networks technology: 1500 octets in Ethernet and 1000 octets are common in X.25 networks. The packet length distribution measured from the real traces exhibits the well-known multi-mode behaviour, with peaks for very short packets and for the different maximum transfer units in the network, with a dominating peak at 1500 bytes, due to the size of Ethernet frame. This specific packet length distribution has a direct impact on the service time and we need a different approach to the queueing analysis.

Discrete-time queueing systems have been a research topic for several decades now and there are many reference works on discrete-time queueing theory. Over the years, different methodologies have been developed to assess the performance of queueing systems. The two main analytical approaches are the matrix analytic method and the transform method for discrete and for continuous-time analyses. Many authors have considered the Geo/G/1 queueing system [Pitts, 2000], [Mirtchev, 2006], [Vicari, 1996], [Zang, 2001].