ROMAI J., **3**, 2(2007), 153–159

QOS ANALYSIS AT THE LEVEL OF NETWORK NODES

Olesia Groza

Moldova State University, Chişinău, Republic of Moldova

olesiagroza@yahoo.com

Keywords: Priority queues, Quality of Service (QoS), Class of Service (CoS), prioritisation, traffic characteristics.

2000 MSC: 90B22, 68M22, 68M10.

1. INTRODUCTION

We might ask ourselves: why is QoS important? The data that are transferred through modern computer networks are becoming more and more diverse. The volume of such multi-type data that is to be transferred and delivered is also growing. Thus, there arises the problem of providing an adequate level of the quality of service (QoS) in such networks. If the resources of a network were unlimited, then all data traffic would be transferred and delivered with a necessary speed without delays and losses. However, the network resources are limited, and the QoS mechanism controls distribution of these according to the needs of the traffic.

Quality of Service is a general concept, which refers to the ability of a network to provide better services for the selected network traffic over various technologies, including Frame Relay, Asynchronous Transfer Mode (ATM), Ethernet and 802.1 networks, SONET and IP-routed networks that may use any or all of these underlying technologies. For example, Ethernet provides two different QoS mechanisms. One mechanism is via 802.1p, which provides eight classes of service. The other mechanisms is via VLANs, whereby traffic can be separated, isolated and prioritised by the VLAN ID. Frame Relay has a QoS mechanism called Discard Eligibility that can be set by the networking device for traffic that can be discarded under network congestion.

2. PROVIDING QOS AND QOS PARAMETERS

The problem of providing QoS in networks was an area of a very active research in the last few years. The first suggested mechanism of providing QoS end-to-end control in a network was called Integrated Service (IntServ). This is a complex of services which guaranteed a corresponding level of quality of service in network data routing and information transferring. A different mechanism appeared later - this has become to be known as Differentiated Service (DiffServ) and it offered different treatment for the network packets depending on their importance. Both mechanisms guarantee QoS with a certain probability.

In order to avoid or diminish agglomeration at a node of a network there are used various algorithms. These algorithms employ mechanisms of prioritisation of the traffic which enters or leaves the node

- FIFO (first-in, first-last),
- LIFO (last-in, first-out),
- PQ (priority queueing),
- CQ (custom queueing),
- RANDOMIZE.

Every node inside the network can be analysed as a queueing system. Thus, QoS control in networks can be reduced to the analysis of the performance characteristics of queueing systems: the mean number of served requests, the number of lost requests, the mean waiting time, the mean service time, busy period, traffic coefficient etc. These characteristics are also called QoS parameters. The difficulty, yet, consists in the lack of existence of an appropriate model for the traffic circulation within the network.

Every queueing system is characterized by certain parameters.

Waiting time of a request represents the time during which the request was waiting its service in the queueing buffer; the *service time* of a request is the time which the server spent for its service. The incoming traffic of a router can exceed its outcoming capacity - therefore, there exists a tampon memory space which is designed for a temporal storage of the packets. Yet, this memory results in the additional delay of the traffic, which add up to the delays related to waiting and servicing. Such delays represent time variation called "jitter". Thus, jitter represents a measure of variation in delays between consecutive packets of a flow. The jitter has the most notable effect in a real time, especially in video and audio applications. Applications of this type expect the packets to arrive at the quite a constant rate with fixed delays between consecutive packets. If the incoming rates vary, then the jitter affects the performance of the application. A certain minimal amount of jitter may be accepted, but if this is increased, then an application may become inapplicable. Some applications, such as audio, for instance, can compensate for small deviations of the "jitter". Thus, VoIP applications reprocess a precedent vocal packet in the case when there is a delay in the next packet's arrival. However, if the next packet is delayed for a long time, then it is simply eliminated when it arrives, which results in a slight audio deformation. All networks introduce some variations of delays (jitter), as the packets are queueing at every node of the network. Yet, if the delays' variations (jitter) are limited, QoS can be maintained. The *total sojourn time* of a request (or, the transfer time) represents the sum of its waiting and service times. This characteristic may have a significant influence on such applications as audio, video, fax transmission etc. Some applications can compensate for small volumes of delays. For instance, VoIP entries provide special local buffers (tampon memories) for delay compensation in the network. Yet, if a certain volume is exceeded, providing a QoS becomes a problem.

Busy period represents a time period during which one ore more stations were busy with servicing of requests or with switching their states. The number of requests in the system (or, system state) represents the number of requests which are in the system at time t ($t > 0$). Yet, another characteristic deserves also a special attention - this is the number of lost requests (lost packets). Due to the fact that the number of places in the waiting line is finite (tampon memory is finite and can be overflown), one can expect losses of the packets when the incoming traffic exceeds the outcoming capacity. Losses can also take place due to mistakes introduced by the physical medium of the transmission. For instance, wireless connections: satellite, mobile network have a high Bits Error rate (BER), which is caused by surrounding medium or by geographical conditions, such as moisture (high humidity), constructions, trees, mountains etc. Losses can also take place when the agglomerated network nodes ignore some packets. Some network protocols, such as TCP (Transmission Control Protocol), offer protection of the packets through retransmission of those ones which have been eliminated from the network. A network can become agglomerated from time to time - more packets are eliminated then and more retransmissions TCP take place then.

It follows that the waiting time influences the number of requests in the system, and the busy period does influence the time of their servicing.

Traffic coefficient, or, relative intensity of the traffic, represents an average number of requests which appear in the system during a service time of one such request. Thus, the service factor is a measure of the system workload. This factor has a big importance because after defining the distribution of the service times all the model characteristics under the study can be expressed as a function of this parameter.

By analysing these parameters one can determine whether the level of offered or received services is realised and how to get insight into a phenomenon of queueing or agglomeration.

There are also other parameters which affect QoS and which cannot be quantitatively measured. However, such parameters provide one with mechanisms of traffic management of network nodes. These parameters are:

- emission priorities;
- discard priorities.

Emission priorities define the order of the flow before it will pass through a node of the network. The traffic of a higher propagation priority is placed in front of the flow of a lower priority. Propagation priority determines also the amount of latency introduced in traffic by network nodes. For instance, the delay of the ”tolerant” applications such as e-mail can be configured so that it will have a lower priority than the delay of ”sensible” applications (such as audio and video applications are). Thus, packets of the tolerant applications can be buffered in the tampon memory until the packets of sensible application are transmitted. Propagation priorities uses a simple transmission mechanism with priority according to which the traffic with a higher priority is always given a preference. A disadvantage of this mechanism is that the traffic with a lower priority might never be transmitted. This may happen if the traffic of a higher priority persists in the network and the network bandwidth is limited.

Discard priorities are used for determining the order in which the traffic leaves network nodes. More specifically, if the network nodes experience agglomeration, then the high priority traffic will leave a node earlier than a low priority traffic. Thus, the departure priority is used to eliminate most eligible traffic first. Traffic with the same QoS requirements can be divided using departure priorities. If, however, a network node is not agglomerated, then the traffic will receive the same treatment (there is no priority). In the absence of the departure priorities it is necessary to divide the traffic in different subflows at the network nodes in order to provide differentiated services. According to the departure priority the packets of different applications are placed in the same waiting line, which, yet, is divided into different virtual subflows, each having different departure priority. For instance, if the packets of the traffic support three departure priorities, then, as a result, three level QoS is provided.

3. APPLICATIONS REQUIRING QOS

Quality of Service is required by different applications, such as the following, to name a few:

- IP telephony or Voice over IP (VoIP);
- video teleconferencing (VTC);
- alarm signalling;
- streaming video on demand;
- streaming audio;
- client/server transactions;
- e-mail;
- file transfer.

Thus, the main objective of studying of queueing systems in the framework of network management is to establish a certain adequate level of the service quality, which depends on the corresponding queueing system performance characteristics: waiting times, service times, the number of lost requests (packets), busy period, traffic coefficient etc.

4. ACKNOWLEDGEMENT

This work was done under support of the SCOPES grant IB7320-110720.

References

[1] *Cisco Systems. Cisco IOS 12.0 Quality of Service*, Cisco Press, Indianapolis, 1999.

[2] Chow, J., *Development of channel models for simulation of wireless systems in OPNET*, Transactions of the Society for Computer Simulation International, **16**, 3(1999), 86-92.

[3] *Nortel Networks, Introduction to quality of service*, White Paper, 2003 (technical document).

[4] Grama, I., Mishkoy, G. *The object oriented programming for queueing systems*, Computer Science Journal of Moldova, **1**, 1(1993).

[5] Green, D. B., Obaidat, M. S., *Modeling and simulation of IEEE 802.11 WLAN mobile ad hoc networks using topology broadcast reverse-path forwarding (TBRPF)*, Computer Communications, **26**(2003), 1741–1746.

[6] Heindl, A., German, R., *Performance modeling of IEEE 802.11 wireless LANs with stochastic Petri nets*, Performance Evaluation, **44**(2001), 139–164.

[7] *Cisco Systems, Internetworking Technology Handbook*, Documentation, `http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.htm`.

[8] Kakaraparthi, R., et al, *Efficient message scheduling in ad hoc networks*, Proc. IEEE Wireless Communications Networking Conf., 2000, 1226–1231.

[9] Peuhkuri, M. *IP Quality of Service*, 1999, `http://www.netlab.tkk.fi/u/puhuri/htyo/Tik-110.551/iwork/iwork.html`.