PRIORITY QUEUEING SYSTEMS AND PRIORITIZATION PHENOMENA IN INFORMATION NETWORKS

[1]Gh. Mishkoy, [2]S. Giordano, [3]A. Bejan,
[1]Academy of Sciences of Moldova, Free International University of Moldova
[2]University of Applied Sciences of Southern Switzerland
[3]State University of Moldova, Heriot-Watt University (Edinburgh, UK)

1. PRIORITIES AND SWITCHING IN QUEUEING SYSTEMS

Queueing systems represent adequate models of everyday phenomena in many aspects of our life. Nowadays, queueing systems theory has a special place in the performance analysis of a broad range of systems in communication theory, logistics and manufacturing.

Priority queueing systems form a large class of queueing systems where the incoming requests are to be distinguished by their importance. Such systems represent adequate models of many aspects of everyday life, when a preferential service is to be granted to certain kinds of requests (demands, customers). Priority queueing systems have also found important applications in the modelling and analysis of computer and communication systems: packets' transfer and routing in computer networks, distributed operations and calculations (multiprocessor OS's, etc), telephone switching systems, mobile phone networks. Some civil services (surgeries, ambulances, fires, etc.) can also be modelled using the concept of priority queueing systems.

The general rule of service in priority queueing systems is as follows: the requests which are in the system and have a higher priority should be served before those that have lower priorities. However, the mode of the server's behaviour in such systems may essentially diversify them. In addition, there are systems in which server needs some time to switch itself from the servicing of one kind of requests to another. All this gives a great variety of the considered systems. Accordingly to these phenomena the description and classification of the priority queueing systems is given in [11].

2. PRIORITY QUEUEING SYSTEMS IN CIVIL APPLICATIONS AND ENGINEERING

Some potential applications for using priority queueing systems with switchover times are discussed here.

2.1. Civil applications: medicine and health care

A bibliography of more than 150 references which consider applications of queueing theory to problems in health care and medicine can be found in [15]. Introduction of priorities is natural in such social sector as health care. The phenomenon of prioritization in queueing processes related to health care and medicine was considered in [8, 10, 13, 16, 17, 19], just few to mention. We stress that the necessity of switchover times – times to switch between servicing of customers of different priorities is natural assumption in all such considerations.

2.2. Applications in engineering and network technologies: QoS and CoS technologies

As an example of engineering application of priority queueing systems with switchover times we consider QoS and CoS network technologies which are modern and important today.

Quality of Service (QoS) and *Class of Service* (CoS) technologies play nowadays a crucial role in the analysis of a network traffic, which is highly diverse and may be characterized in terms of *bandwidth*, *delay*, *loss* and *availability*. Some more specific characteristics can also be considered.

Most of the network traffic is IP-based today. On the one hand it is beneficial, as it provides a single transport protocol and it simplifies maintaining of the hardware and software products. However, IP-based technologies have some drawbacks. First off all, under the IP protocol network packets are delivered through the network without taking any specific path. This results in the unpredictability of the quality of service in such networks.

However, today networks deal with so many types of traffic that these may interact in a very unfavourable manner while being transmitted through the network. QoS and CoS technologies serve to ensure that diverse applications can be properly supported in an IP-network [12]. This is achieved by distinguishing between different types of data and by managing them using the mechanisms of data prioritization.

We introduced in [11] a diverse class of priority queueing systems involving switching to describe, model and analyze phenomena which involve prioritized queueing and may take place in a studied or designed network. We suggested that some performance characteristics of such priority queueing systems can be used for estimating and providing a respective Quality of Service.

In the following exposition we discuss briefly the QoS and CoS methodologies and their applications in analyzing and modelling networks. Then, we consider an example of usage of such systems, and, in particular, we discuss the benefits of using them for obtaining QoS in WLANs.

3. QOS AND COS METHODOLOGY IN NETWORK TRAFFIC ANALYSIS AND MANAGEMENT

Quality of Service is a general concept referring to the capability of a network to provide better service to selected network traffic over various technologies, including *Frame Relay*, *Asynchronous Transfer Mode* (ATM), *Ethernet* and *802.1x networks*, *SONET*, and IP-routed networks that may use any or all of these underlying technologies [5].

Define by a *flow* in a broad sense a combination of packets passing through a network. Basically, QoS enables to provide in a network a better service to certain flows by assigning the higher priority of a flow or limiting the priority of another. This can be done in different ways: mostly by designing corresponding queue management mechanisms.

One can represent the basic QoS architecture by the following three components and steps [5]:

- QoS marking techniques for coordinating QoS from end-to-end between network elements
- QoS within a single network element (e.g. queueing, scheduling, traffic-shaping tools)
- QoS policy, management and accounting functions to control and administer end-to-end traffic across a network.

We refer in this paper mostly to a QoS of a single network element (i.e. to a second step of the QoS providing architecture scheme given above).

QoS within a single network element, or node, can be specified by a *congestion management*, *queue management*, *link efficiency*, and shaping/policing tools.

The Class of Service concept is a concept of the flow network traffic division into different classes. This concept provides class-dependent service to each packet in a flow, depending on which priority class it does belong to [14]. CoS provides end-to-end prioritization for frame relay and ATM traffic over IP networks. In a framework of CoS traffic is prioritized by setting the *Differentiated Services* code in the header of an IP data packet.

4. PRIORITIZATION IN INFORMATION NETWORKS

Prioritization plays the crucial role in QoS and CoS technologies.

In information networks it is desirable to provide shorter waiting times for control packets (packets that contain information about network status), voice connection packets and packets associated with messages which should be delivered urgently.

There are many ways to attribute preferences. However, on a conceptual level, there are not so many ways to provide preferential service in a queueing system or queueing network [2]. In [11] a wide range of service disciplines in priority queueing systems involving switching between flows has been described.

For examples and more account on prioritization and its forms the reader is referred to [2]. Description of some queueing disciplines implemented at nodes of an ad-hoc network can be found in [9]. QoS in ad-hoc networks and mechanisms of data prioritization in such networks are discussed in [1, 18] and references therein.

One can expect that Quality Design (QD) will be fundamental also for the new emerging mobile networks such as opportunistic networks or Mesh networks. Even very best effort networks, as opportunistic networks [7], or networks developed for the daily use, as mesh networks [3], will be faced with priority issues.

5. DISCUSSION AND EXAMPLE

The mentioned description of the priority queueing models given in [11] can be considered to describing the processes taking place in communication and information networks. As it has already been pointed out, in order to design better communication network and to provide higher level of quality of service, it is really important to be able to evaluate network performance parameters. One should pay special attention on the analysis of the node traffic characteristic.

However, modern communication networks and network node processes can be much more complicated. Particularly, node incoming flows may not be Poisson and may be described by a much more complicated distribution. As discussed above the input flows may exhibit properties of *burstiness*, *correlation*, *self-similarity*. In such cases development of traffic models is more sophisticated and analytical methods became less powerful. The assumption about non-Poisson nature of arrival flows makes analytical methods to be less efficient in providing information on the system performance characteristics.

The Java package *PQSST* presented in [4] allows to obtain full chronology of the system under study. Additionally, it provides summary on busy periods statistics, idle periods statistics, mean waiting times of requests, etc.

It is believed that the package *PQSST* will be of real interest for the performance analysis of priority queueing systems with switchover times, and particularly, in the context of QoS provision in communication traffic systems.

We continue with an example of usage of the described systems. This example is based on a Cisco Priority Queueing technology which is described in [6].

Priority queueing is useful for making sure that mission-critical traffic traversing various WAN links gets priority treatment. For example, Cisco uses priority queueing to ensure that important Oracle-based sales reporting data gets to its destination ahead of other, less-critical traffic. Priority queueing uses static configuration mechanism and does not automatically adapt to changing network requirements. In this example prioritization represents the process of placing data into four levels of queues: high, medium, normal and low. This is shown schematically in Figure 1.
It is easy to see that this process of prioritization can be modelled as $G_4 | G_4 | 1$ priority queueing system with postponable priority service discipline and correspondingly chosen densities $a(t)$ and $b(t)$ of interarrival and service times, respectively (see [11]). The discipline of switching can also be appropriately chosen.

However, one may choose to consider a service discipline other than non-preemptive one (as *postponable service discipline* is, accordingly to the classification given in the previous section) in order to minimize mean waiting times of the packets, for instance. The package PQSST may be useful for these purposes, unless the incoming flows are of Poisson type (analytical methods may be applied then).

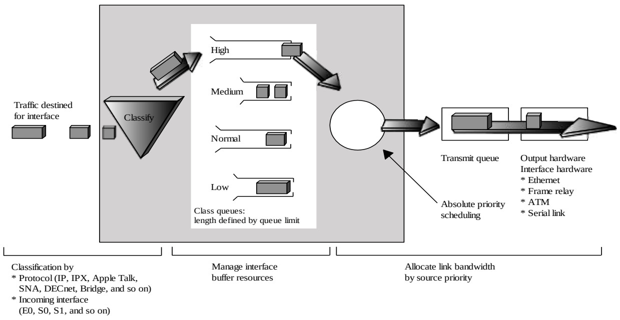

Figure 1. Priority Queueing Places Data into Four Levels of Queues (reproduced from CISCO documentation [6]).

One of the most important characteristics of the priority queueing systems is the node traffic coefficient. This quantity plays the crucial role in estimating node QoS. The role of stationarity conditions expressed with the help of traffic coefficient is very important on the way of providing network QoS. The further work will establish quantitative relationships between this characteristic of a queueing system and characteristics involved in estimation of a network QoS.

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

References

1. Aad, I., Castelluccia, C. *Priorities in WLANs. Computer Networks, 41, pp.505-526, 2003.*

2. Anderson, R.H. et al. *Securing the U.S. Defence Information Structure. Rand Monograph Report, 2003.*

3. Bahl, V. *MESH networking.* To appear in *IEEE Communication Magazine, 2007.*

4. Botezatu, M., Bejan, A. *Priority Queueing Systems with Switchover Times: Modeling and Analysis in Java. The XIV Conference on Applied and Industrial Mathematics, Satellite Conference of ICM 2006, Chişinău, pp. 49-52, 2006*

5. *Cisco Systems. Cisco IOS 12.0 Quality of Service. Indianapolis: Cisco Press, 1999.*

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

7. Conti, M., Giordano, S., *MANETS: the real work.* To appear in *IEEE Communications Magazine, 2007.*

8. Hannan, E.L., Gimborne, C.J. *Predicting the impact of instituting a priority readmission policy in nursing homes. Computers and Operational Research 14, pp. 493-505, 1987.*
9. Kakaraparthi, R., et al. *Efficient message scheduling in ad-hoc networks. Proc. IEEE Wireless Communications Networking Conf., pp.1226-1231, 2000.*
10. Kapadia, A.S., Chiang, Y.K., Kazmi, M.F. *Finite-capacity priority queues with potential health applications. Computers and Operational Research 12, pp. 411-420, 1985.*
11. Mishkoy, Gh., Giordano, S., Andronati, N., Bejan, A. *Priority Queueing Systems with Switchover Times: Generalized Models for QoS and CoS Network Technologies and Analysis. Technical report, 2006.* *http://www.vitrum.md/andrew/PQSST.pdf*
12. *Nortel Networks. Introduction to Quality of Service. Technical document, 2003.*
13. Panayiotopoulus, J.C. *Solving queueing systems with increasing priority numbers. J. Operational Research Society 31, pp. 637-646, 1980.*
14. Peuhkuri, M, *IP Quality of Service. 1999. http://www.netlab.tkk.fi/u/puhuri/htyo/Tik-110.551/iwork/iwork.html*
15. Preater, J. *A bibliography of queues in health and medicine. Bibliography, 2001. http://www.chcm.ubc.ca/comm682/QueueingBib.pdf*
16. Rege, K.M., Sengupta, B. *A priority-based admission scheme for a multiclass queueing system. AT&T Technical J., pp.1731-1753, 1985.*
17. Shimshak, D.G., Gropp Damico, D., Burden, H.D. *A priority queueing model of a hospital pharmacy unit. European J. of Operational Research 7, pp. 350-354, 1981.*
18. Stine, J.A., Gustavo de Veciana. *A paradigm for Quality of Service in Wireless Ad hoc Networks Using Synchronous Signaling and Node States. IEEE Journal on Selected Areas in Communications. 22(7), pp. 1301-1321, 2004.*
19. Taylor, I.D.S., Templeton, J.G.C. *Waiting time in a multi-server cutoff-priority queue, and its application to an urban ambulance service. Operations Research 28, pp. 1168-1188, 1979.*

Recomandat spre publicare: 24.11.2006