Translation from Ukrainian into English

УДК 621.311

O.V.Tymchenko[1) 2)], Doctor of Eng. Sc., Professor,
Sami Askar [2)], Ph.D. candidate, Alhihi Muhamad[2)], Ph.D. candidate, Al-bdur Nashat [2)], Ph.D. candidate

INVESTIGATION OF MECHANISMS OF SERVICE QUALITY CONTROL IN MULTISERVICE NETWORKS

The mechanisms of maintaining a specified quality of service (QoS) for telecommunications networks and determining time-response characteristics are examined for the square root criterion.

Introduction

Multi-service telecommunications network is a unified telecommunication infrastructure for transfer, switching of free-hand type traffic created by interacting of users and providers of telecommunications services with controlled and guaranteed traffic parameters. Such networks should guarantee a certain quality of provided connections and services. Multi-service network traffic is variable because it includes multimedia applications, file transfer, voice transmission, videoconferencing, etc.

It is classified according to the following identifiers:

– network identity (local or global);
– traffic type – depends on applications using it;
– critical time delay or non-critical one;
– segmented, squeezed or on-line, etc..

Table 1 shows various types of traffic transferred in real networks and its characteristics.

[1] Wyższa Szkoła Humanistyczno – Ekonomiczna, Zakład Metod Przetwarzania Informacji, ul. Rewolucji 1905r. nr 64, 90-222 , Łódż, Poland

[2] Національний університет „Львівська політехніка”, Інститут телекомунікацій, радіоелектроніки та електронної техніки, каф. Телекомунікацій, вул.С.Бандери, 12, 79013, Львів, Україна

Table 1
Traffic types and characteristics

Service	Users' class	Maximum transmission rate	Bursting ability	Peak connect time	
				Tp, s	Tc, s
Telephony	Private sector	64 kbit/s	1	100	100
	Business sector	64 kbit/s	1	100	100
	PABX	64 kbit/s	1	100	100
Fax	Business sector	2 Mbit/s	1	3	3
	PABX	2 Mbit/s	1	3	3
Data communications	Business sector	2 Mbit/s	1	1	1
	PABX	2 Mbit/s	1	1	1
Video telephony	Private sector	10 Mbit/s	5	1	100
	Business sector	10 Mbit/s	5	1	100
	PABX	10 Mbit/s	5	1	100
Video search	Private sector	10 Mbit/s	54	10	540
	Business sector	10 Mbit/s	18	10	180
	PABX	10 Mbit/s	18	10	180
	Service center	10 Mbit/s	48	10	480
Document retrieval	Private sector	64 kbit/s	200	25	300
	Business sector	64 kbit/s	200	25	300
	PABX	64 kbit/s	200	25	300
	Service center	64 kbit/s	200	25	300
Data on call	Business sector	64 kbit/s	200	0.04	30
	PABX	64 kbit/s	200	0.04	30

As we can see, there are every possible traffic types in multi-service network. Different demands as to transfer rate, bursting ability, connect time, etc. are made on each of these type.

For instance, if it is necessary to make connection on the basis of telephony, the critical needs for such a session are to be transfer rate and communication delay in channel, whereas the parameter of guaranteed error-free delivery can be ignored. In case of data transfer, the main criterion is to be the adequacy of data received and the safety (protection from data diddling, authenticity).

Therefore, in case of simultaneous transfer of different traffic types a situation of channel capacity shortage can take place.

In this case, to provide a quality transfer it is necessary to use traffic control algorithms, because there exists the possibility to loss valuable information and the impossibility to carry out subsequent transfer owing to full channel occupancy.

Traffic prioritization mechanisms

When processing the whole of traffic with a similar priority it is impossible to maintain a high quality of service (QoS) for different applications with a limited channel carrying capacity. Some network traffic such as video on demand (VoD) is designed for a maximum allowed delay and a high carrying capacity. However, if such traffic has the same priority as the other, delay can be unpredictable and carrying capacity can be insufficient owing to that other applications will use all network resources. In this connection, the multi-service network should be created so that it could guarantee a necessary level of services for each application.

The applied Resource Allocation Fairness Method is one of basic principle of network operation on the basis of IP protocol. The main idea consists in an equivalent allocation of bandwidth among competing batches/flows irrespectively of loading. In the general case, when the network loading is low, service standards for all batches/flows are met. However, when network point loading becomes high, there is an increased probability of that batches of one flow will use all network resources blocking all other flows.

To maintain a specified quality of services, the methods of applications traffic allocation according to categories and prioritizing of individual flows are widely used, thereby a traffic with a high priority receives necessary network resources according to enquiry characteristics irrespectively of requirements as to traffic capacity of less important applications.

When transferring each data type (IP-telephony (VoIP), video conferencing, VoD, etc.) different requirements are put forward [3], which could be laid down as restrictions on delivery time for batches of different types. Let us single out two types of such restrictions:

1) probablistic restriction (assignment of allowable probability $p_{i\text{доп}}$ of the excess of specified restriction $t_{i\text{доп}}$ for batch delay time t_i in multi-service network):

$$P\ (t_i > t_{i\text{доп}}) < p_{i\text{доп}},\ i=1\ldots n, \quad (1)$$

2) average restriction (restriction $t_{i\text{доп}}$ for average delay time $t_{i\text{сер}}$):

$$t_{i\text{сер}} < t_{i\text{доп}},\ i=1\ldots n, \quad (2)$$

where n – number of traffic types in network.

The indicated requirements can be met by using control methods providing an effective allocation of channel capacity among different traffic types. Optimum Batch Priority Allocation is one of such methods. To evaluate the effectiveness of prioritizing critical batches as to delay, models with heterogeneous inflow of orders are used that allow an analysis of properties of systems with a priority level and the formulation of recommendations for engineering of such networks, for instance, by the determination of necessary channels capacity.

ToS mechanism

The application of Optimum Priority Allocation Method is reasonable and possible provided that switches and routers can differentiate batches with different priorities. The solving of problem of maintaining QoS was provided for already at the development stage of IPv4 protocol specification. In headers of IP-batchs a special ToS (Type of Service) bite is used which would be used for instruction on necessary quality of service. Currently, ToS bite is widely used for the traffic differentiation according to requirements as to QoS. ToS bite contains three priority bites (IP-priority) and four ToS bites (one bite is not in use) (Fig.1).

Fig. 1. Datagram format

Version Hlen Service type Overall length
Identifier Toggles Fragment designator
Lifetime Protocol Heading hash total
Sender's IP-address
Receiver's IP-address
IP-options Filled in by
Data

Three priority bits allow establishing of eight priority levels:

111 – network control;
110 – internetwork control;
101 – CRITIC/ECP;
100 –flash overrider
011 – flash;
010 –immediate;
001 –priority;
000 –routine.

Four bits digitize service type:

0000 – normal service;
0001 –minimize monetary cost;
0010 –maximum reliability;
0100 – maximum capacity;

1000 – minimum delay.

One bit only of four in TOS can take the value of 1. Default values are equal to zero. The majority of recommendations are evident. Thus, in telnet the response time is the most important, and the reliability for SNMP (network control).

Fig. 2. ToS field format

Priority Not in use

ToS field format is specified in document RFC-1349. Bites C, D, T and R define a request for the datagram delivery method. Value D=1 requires a minimum delay, T=1 – a high capacity, R=1 – a high reliability, a C=1 – low cost. ToS plays a key role in routing of batches. Internet does not ensure a requested ToS, but many routers take these queries into account when routing (OSPF and IGRP protocols). Table 2 gives recommended ToS values for various protocols.

Table 2

ToS values for different protocols

Procedure	Minimum delay	Maximum communications capacity	Maximum fidelity	Minimum value	TOS code
FTP control, FTP data	1	0	0	0	0x10
	0	1	0	0	0x08
TFTP	1	0	0	0	0x10
DNS, UDP, TCP	1	0	0	0	0x00
	0	0	0	0	0x10
	0	0	0	0	0x00
telnet	1	0	0	0	0x10
ICMP	0	0	0	0	0x00
IGP	0	0	1	0	0x04
SMTP control, SMTP data	1	0	0	0	0x10
	0	1	0	0	0x08
SNMP	0	0	1	0	0x04
NNTP	0	0	0	1	0x02

ToS mechanism plays a key role in batch routing. Although ToS is not guaranteed when transmitted by Internet network, however, a substantial number of routers take these queries when selecting a route (OSPF and IGRP protocols). as As a rule, before mid-90-ies ToS field was ignored.

However, with an increased demand for maintaining quality of service a demand for methods of the implementation. In subsequent engineering designs (RFC-2474, Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers) ToS field was replaced by field (Differentiated Services Code Point), where six low bits are chosen for DS (Differentiated Services) code, and two high bits are reserved. Six bits define service class (Fig.2).

DS5	DS4	DS3	DS2	DS1	DS0	CU	CU

Fig. 2. DSCP field format

In Fig.2 bits DS0-DS5 define the class selector. Values of these code are given in Table below. The DSCP standard default value по замовчуванню - 000000.

Class selector	DSCP
Priority 1	1000
Priority 2	10000
Priority 3	11000
Priority 4	100000
Priority 5	101000
Priority 6	110000
Priority 7	111000

PHB technology

On the basis of DSCP " per Hop Behavior " (PHB) technology was developed. To provide traffic servicing, a client concludes with provider a Service Level Agreement (SLA) on providing a certain service level. SLA can contain traffic condition rules, fully or partially – a part of Traffic Condition Agreement (TCA). TCA defines the rules for classifier and corresponding traffic profiles as well as the rules for marking, batching and batch rejecting to be in use for any flows. Within the framework of this policy DSCP codes inside classes are defined. At the boundaries of a domain the traffic conditioning takes place, i.e. a traffic classification providing the analysis of input batches, a collocation of acquired information with flow table as well as batch marking with a special code point DSCP (DiffServ Code Point). For instance, for the policy of immediate EF diversion the value of DSCP=101110 is recommended. This policy correspond to a highest level of services.

These functions are performed by so called domain port-access. the port functions are as follows:

1. input traffic analysis (title retrieval of L3 and L4);

2. conformity check in the flow table of commutator switch and batch allocation one-by-one according to flow description;
3. filtering of non-classified traffic (batches that do not belong to any logical flow);
4. set communication rate restriction for each flow ("token bucket" algorithm);
5. marking of batch IP-header in TOS field by code point DSCP (DS Code Point).

The sequence of batches each of which can be unambiguously identified on 16-bite combination of first 64 bites of IP-header and/or TCP/UDP header (notification port number) should be understood by flow.

Subsequently, the traffic processing in intermediate centers, the approval of decision on batch routing in a certain queue is carried out exclusively on code point DSCP placed in IP batch header (TOS field). The processing of classified traffic inside domain is carried out at commutation rate – it is sufficiently to read out 6 bits of code word and send the batch to a corresponding queue, whereupon the “weighed fair services” algorithm enters in operation (Fig.3). Interior port connects two objects in DiffServ domain. For instance, these are to be trunk ports of Gigabit Ethernet switches connected through fiber optic. The port functions:
1. inbound traffic analysis (DSCP readout)
2. batch allocation in queues according to DCSP
3. reassignment of code word in case the port is an output from DiffServ domain to provide matching of QoS levels among switches of different manufacturers.

Exterior port – connects DiffServ domain to outside (domain top). The port functions are similar to access port functions. The exterior port processes traffic entering a domain.

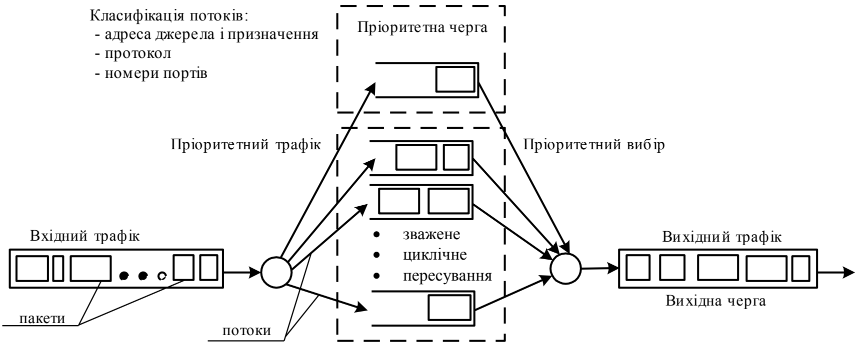

Fig. 3.Mechanism of weighed fair services

Flow classification: Priority queue
- source address and assignment
- protocol
- port numbers

-

Priority traffic — Priority selection

Inbound traffic — Outbound traffic

- weighed
- cyclic
- movement

Batches — Ouput queue

Flows

In Cisco routers, two low bits of three are used to classify batches and select an event queue. On default, 10% of band are assigned for 0 class and 20%, 30% and 40%, respectively, for classes 1, 2 and 3. For queues on the basis of QoS classes, the batches that do not pertain to any group are related to group 0 and automatically have 1% of total communications capacity of a group. A total weight of remaining groups can not exceed 99%. If an unallocated band is available, it relates to 0 group. In protocol, IPv6 field has a priority of 4 bits. Bits C, D, T and R characterize a desirable method of datagram delivery. Table 1 gives standardized values of TOS field of IP-batch.

Procedure for determination of main network parameters

To evaluate the efficiency of priority methods for traffic control in mult-service networks, a queueing system with non-uniform flow of batches of *n* types that enter a channel with intensities of X1,…,Xn is used as a basic transmission channel model.

When solving problems of allocation of network resources among different communication services, the user of each service is characterized - on the one hand – by the traditional traffic parameters:

- intensity of input flow of calls for providing services by k-th communication service $\eta^{(\kappa)}$, calls/year;
- mean communication session time $T_{\tilde{n}}^{(\kappa)}$, s;
- specific calling rate $\gamma^{(\kappa)}$, Earl;

- on the other hand – by parameters of random process, but by such ones that characterize each user:

- peak (maximum) bit transfer rate $B^{(\kappa)}_{max}$, bit/s;
- mean bit transfer rate B_{cep}; bit/s;
- batching ability $K_{п}$ that is determined by the relationship $B^{(\kappa)}_{max}/B_{cep}$; bit/s;
- mean peak time $T_p^{(\kappa)}$,s.

Assume the following restrictions:

- number of calls for providing services $\eta^{(\kappa)}$, calls/year.;
- peak transfer rate $B^{(\kappa)}_{max}$;
- network calling rate coefficient ρ;
- mean message length μ.

Network communication capacity is determined from the formula:

$$C = \frac{C_{min}}{\rho} . \tag{3}$$

where C_{min} – minimum communication capacity:

$$C_{min} = \frac{1}{\mu} \cdot \sum_{i} \lambda_{i} . \tag{4}$$

Subsequently, main relationships for determining of time-response characteristics for the square root criterion which optimizes traffic-handling capacity on minimum weighed mean time of batch transfer over network are given.

Communications capacity of i -th channel [bit/s]:

$$C_{i\text{опт}} = \frac{\lambda_{i}}{\mu} + \frac{C \cdot (1-\rho) \cdot \sqrt{\lambda_{i}}}{\sum_{j} \lambda_{j}} , \tag{5}$$

where λ – batch intensity in a channel.

Batch delay time in i-th channel [c]:

$$\overline{T}_{i} = \frac{1}{\mu \cdot C_{i} - \lambda_{i}} . \tag{6}$$

Minimum batch delay time in a network:

$$\overline{T}_{min} = \frac{\sum_{i} \lambda_{i} \cdot \overline{T}_{i}}{\gamma} , \tag{7}$$

where γ – a total batch entry intensity in a network.

In the event if batches of one class have an equal length and all the flows are the simplest, the mean delay of batch of *i*-th type when using the traffic control method on the basis of relative priorities is determined from the formula:

$$t_{i} = \frac{\sum_{j=1}^{n} \lambda_{j} L_{j}^{2}}{(V - \sum_{j=1}^{i-1} \lambda_{i} L_{i}^{2})(V - \sum_{j=1}^{i} \lambda_{i} L_{i})} + \frac{L_{i}}{C} , \; i=1 \ldots n, \tag{8}$$

where L_i – average *i*-type batch length.

Conclusions

Methods of maintaining a specified quality of service (time-response characteristics in multi-service networks by means of the mechanism of weighed fair services and determining time-response characteristics for the square root criterion which optimizes traffic-handling capacity on minimum weighed mean time of batch transfer over network are considered.

1. *Õîëë Ý*. Приоритеризация трафика в сетях IP // Сети и системы связи. 1988. №11 (33).
2. *Òèì÷åíêî Î.Â., ÊèðèêÌ .I*. Моделі системи проектування трафіку та контролю якості послуг в мультисервісній мережі // Комп`ютерні технології друкарства. Зб. наук. пр. – Вип. 17. – Львів: УАД. – 2007. – 260 с. – С.301-310.
3. *Òèì÷åíêî Î.Â., Âåðõîëà ÁÌ., Ñàì i Àñêàð.* Критерії ефективності функціонування комп'ютерних мереж // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Вип.43. – К.: 2007. – С.184-190.
1. Holl E. Traffic prioritization in IP networks// Networks and telecommunications systems. 1988, No. 11 (33)
2. Tymchenko O.V., Kirik M.I. Models of system for traffic design and maintaining quality of services in multi-service network// Printing computer-aided technologies. Lviv, 2007 – P. 260 – P. 301-310
3. Tymchenko O.V., Verkhola B.M., Sami Askar. Criteria of performance efficiency of computer-aided networks//Simulation and IT. NAS of Ukraine, Kyiv 2007, P. 184-190.

Certification of Accuracy

This is to certify that this is a true and correct English translation of the attached document by certified translator ***Zhigovskaya Anna****, "Big Ben" Translation Bureau, Mykolayiv City, Ukraine, ph. +38 (0512) 477321, www.bigben.mykolayiv.com*