ARIJANA MACURA, M. Sc.
E-mail: arijana.macura@fpz.hr
DRAŽEN KOVAČEVIĆ, D. Sc.
E-mail: drazen. kovacevic@fpz. hr
University of Zagreb
Faculty of Transport and Traffic Engineering
Vukelićeva 4, 10000 Zagreb, Republic of Croatia
IGOR FRANOLIĆ, B. Eng.
Bolfan 44, 42230 Ludbreg, Republic of Croatia

Technology and Management of Traffic
Review
U. D. C.: 656 : 621.397.442 : 681.3
Accepted: Jan. 24, 2005
Approved: May 3, 2005

INTEGRATION OF MULTIMEDIA TRAFFIC INTO COMPUTER NETWORK

ABSTRACT

In the recent several years there has come to a rapid increase in the multimedia traffic in computer communication networks. New applications of communication networks have set new, high requirements on the networks, due to the need to provide the users with high-quality service of multimedia data transmission. In response to such requirements, numerous mechanisms have been developed and they insure the service of prioritizing traffic in the communication networks which results in the possibility of determining the quality of service necessary for individual network applications.

KEY WORDS

multimedia traffic, communication networks, TCP/IP protocol, TCP transmission algorithms, data flows

1. INTRODUCTION

This work describes the technologies of network traffic prioritizing used currently in the computer communications, and considering especially the technologies that will be accessible with the operation system Microsoft Windows 2000. The focus is on the Windows 2000 operation system because the priority allocation, i. e. quality allocation of the communication network transmission service has the right effect only if realized from the beginning to the end of the communication connection, which means from one to another computer which hold the applications that require communication. Therefore, it is necessary for the priority allocation protocol to be supported by end computers and the entire communication equipment used for transmission.

2. QUALITY OF SERVICE IN COMPUTER NETWORK

The quality of service is the capability of the network to provide better service to particular types of network traffic. This capability can be realized on different network technologies (e. g. Frame Relay, ATM etc.), as well as at the level of different network protocols (most often TCP/IP protocols).

The quality of service in computer network transmission is the capability of communication devices and transmission media to realize the minimal requirements, i. e. the transmission quality required by a certain network application. The realization of this function requires special control mechanism of the network traffic, determining of the transmission quality required, and the right of an application or user to actually receive such transmission quality. Of course, the basic pre-condition is for the communication network to actually be able to realize such transmission quality. The function of the quality of service in computer networks must also satisfy the end users, but so that the requirements of all the different applications used in the given network environment are also met.

The basic indicators that realize better service in the network traffic include:

- support to assign and reserve throughput,
- reduction of traffic loss during transmission,
- avoiding or managing congestions of single network segments,
- designing of network traffic,
- setting the traffic priority across the entire network.

Figure 1 shows the relation of the quality of service in transmitting data in ATM network architecture

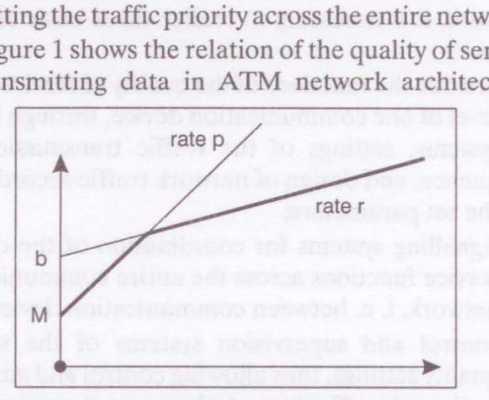

Figure 1 - Example of relation of parameters regarding the transmission service quality in ATM network architecture

with the support for the quality of service. This relation is called also the transmission curve, and defines a group of parameters that determine the specification of traffic flow (*T-SPEC or traffic specification*). In ATM networks these parameters can be expressed as the transmission function of values:

$$A(t) = \min(M+pt, rt+b)$$

where single parameters are:

M – maximal value of network frames on the given segment,

p – maximal throughput of the network segment,

b – network tolerance to peak loading above maximal throughput,

r – sustainable network throughput.

Regardless of the size or purpose of the communication network, all networks can realize certain advantages of using the functions of the quality of service. Of course, different types of users and purposes have also different requirements on the level of service. Business networks require secure transmission and allocation of priorities to critical business applications, in an environment in which, apart from business applications today e-mail and WWW multimedia traffic are very intensively transmitted as well. The companies, the Internet access providers require maximal reliability of the system, and the possibility of controlling the network capacity used by the users (guaranteed speed that has to be realized by the user). These companies also very often enter the transmission area of telephony and other multimedia communications (videoconferences), and therefore have great need for the functions of the quality of service.

Small users and companies want to have functions of the quality of service in order to maximally use the available throughput of their own WAN and Internet connections.

2.1 Architecture of the quality of service

In order to provide the quality of service to network traffic at the level of the entire communication network, it is necessary to realize three basic components:

- provide the functions of the quality of service at the level of one communication device, through buffer systems, settings of the traffic transmission sequence, and design of network traffic according to the set parameters;
- signalling systems for coordination of the quality service functions across the entire communication network, i. e. between communication devices;
- control and supervision systems of the service quality settings, thus allowing control and administration of traffic through the computer network.

It should also be noted that not all the functions and technologies are applicable to all the communication devices. For instance, routers at the edges of WAN network do not perform the same operations as the routers that form the backbone of that same WAN network, and therefore the service quality functions they perform can be different as well. In order to adapt the network for transmission, e. g. IP telephony, it is necessary to know the functions of all the communication devices in the network, and then to adjust adequately the service quality functions. One of the main rules that can be applied to the routers is the rule of classification according to the function to edge and central network devices. Edge routers perform:

- classification of network traffic,
- access control,
- configuration management.

Central routers perform:

- congestion management,
- avoiding congestion of network segments.

2.2. Parameters of the quality of service

Different applications have different requirements regarding the transmission method of their data. Applications mainly expect the network to be capable of transmitting data at the rate at which they are generated. Besides, the applications show also the sensitivity to various other transmission parameters, such as delay and delay variations in transmission. Also, some applications are more sensitive than others to the problem regarding loss of packets, i. e. parts of data during transmission. These characteristics of network applications are expressed through the following parameters that define the required transmission service quality:

- *Throughput capacity* – the rate at which the data generated by the application have to be transmitted on the computer network;
- *Delay* – the duration of transmission of every datum that can be tolerated by the application;
- *Jitter* – maximal delay variation that can be tolerated;
- *Losses* – percentage of data lost in transmission that are allowed.

The mechanism of priority allocation and service quality allocation must be insured in order to meet the communication requirements of a certain application, i. e. to reserve the free capacity of the computer network to the extent which satisfies the requirements.

2.3 Basic mechanisms of traffic control

The computer network connects the end users by a complex system of communication devices of various purposes and characteristics. Each of these systems is connected with other devices by communication media (cables, radio transmission, etc.). Regarding various forms of network topology, various types and capacities of devices and transmission media, it is logical that traffic congestions may occur in some network devices. The devices solve such situations by memory buffers which store the data in transmission until the congestion has passed. Depending on the fact whether the network devices can temporarily store data in case of congestion, this results in the delay and delay variation in transmission or to data loss. The capacity, i. e. rate at which the network devices can forward data, and the capacity of the memory buffers for temporary storage in case of congestion, represent the basic resources which perform the allocation of priorities and the allocation of the service quality in communication.

For instance, in case of congestion, the network device can intelligently decide to send first the traffic which is less sensitive to delay, and store into the buffers the traffic which is more sensitive to delay. In this way, the resource of speed, that is the device capacity or throughput is allocated primarily to delay-sensitive traffic, whereas the delay-insensitive traffic is allocated to the buffer. In order to be able to define the data traffic as sensitive or insensitive to a certain transmission quality parameter, and thus to be allocated to a certain device resource, it is necessary for the device to have a kind of the traffic classification mechanism, i. e. data flow classification mechanism. This is achieved by the device detecting separated data flows, according to the addresses of the conversation participants and type of traffic, and allocating such flows to special queues within the device (e. g. 1st queue for critically sensitive traffic, 2nd queue for medium sensitive traffic, etc.). The transmission data queues are then served according to the defined algorithms of serving each queue, i. e. traffic type. Thus, for the proper functions of allocating the traffic priority, it is necessary to have two separate mechanisms in the device:

- identification of different data flows through the device and their classification per type;
- guidelines on how each of the data flow types has to be forwarded through the device.

These two mechanisms together are called communication traffic control mechanisms

3. TECHNICAL APPROACHES TO SOLVING THE PROBLEM OF MODERN ROUTING

Since due to their capabilities the routers will be indispensable in further development of advanced computer networks, the attempts are made to avoid either partly or completely their disadvantages regarding their performance. There are several technical solutions that offer improvement of the routing performances. Further in the text the best known of these solutions are described.

3.1. Minimizing of routing requirements

The basic idea behind this approach is the known rule "switching wherever possible, routing only where necessary". The idea is to completely eliminate or at least reduce the routing to a minimum. The usual planned method of carrying out this solution is to replace the router by switches wherever possible. Still, the standard *Layer 2* switches have the following limi-

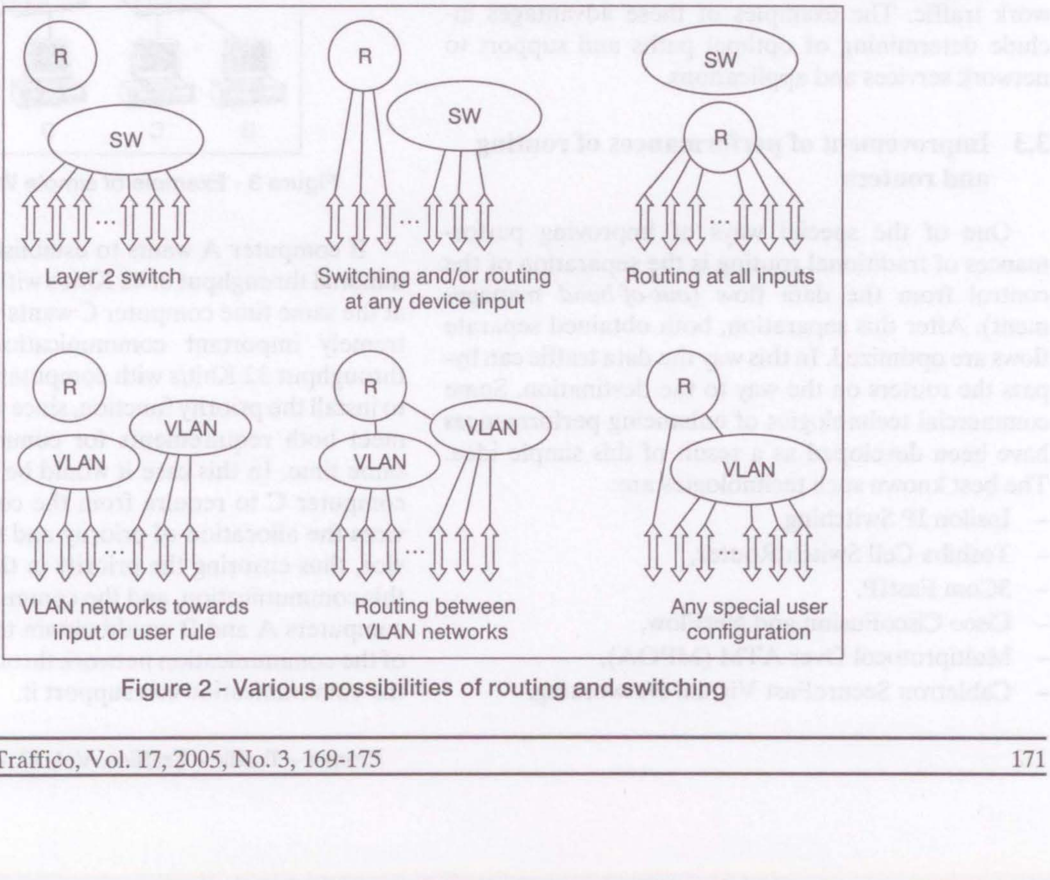

Figure 2 - Various possibilities of routing and switching

tations, which make it impossible to use them instead of routers in all situations:
- limited possibilities of upgrading functionality in order to increase the scalability of the switching part of the network;
- limited possibilities of installing support for new network protocols and technologies;
- limited possibilities of supervision and control,
- need to input big changes in the network operation method, in order to achieve the expected results.

One of the often used options is the use of some of the new network technologies, such as ATM, which use exclusively switching in forwarding network frames. Still, in that case complete replacement of the existing network infrastructure is necessary.

3.2. Integration of routing and switching

Integration of the switching and routing functions in the same device can give excellent results regarding performances, and easy administration and integration into the existing networks, if performed in the right way. One of the possibilities of performing this integration is to add router program support to standard switches. However, this approach has two key drawbacks. First, it does not make the network administration easier, because it only inputs additional routers into the existing infrastructure. Second, this approach does not make full use of the possibilities offered by the synergy between routing and switching. The real advantages of routing and switching integration can be achieved only through the functions of optimization and taking over the advantages of both technologies, according to the current need of the network traffic. The examples of these advantages include determining of optimal paths and support to network services and applications.

3.3 Improvement of performances of routing and routers

One of the special ways of improving performances of traditional routing is the separation of the control from the data flow (*out-of-band* management). After this separation, both obtained separate flows are optimized. In this way the data traffic can bypass the routers on the way to the destination. Some commercial technologies of enhancing performances have been developed as a result of this simple idea. The best known such technologies are:
- Ipsilon IP Switching,
- Toshiba Cell Switch Router,
- 3Com FastIP,
- Cisco CiscoFusion and NetFlow,
- Multiprotocol Over ATM (MPOA),
- Cabletron SecureFast Virtual Networking.

Due to complexity and difficulties that occur in defining the feasible solutions, "route once, switch several times", communication industry is developing alternative solutions that attempt to enhance the performances of the traditional routers, without changing their operation method and implementation on the network. The products that resulted from these attempts are called *multi-gigabyte routers*, and several companies have already developed these (Cisco, Packet Engines, Juniper, etc.).

4. INTEGRATION OF PRIORITY ALLOCATION FUNCTIONS INTO ADVANCED COMPUTER NETWORKS

Advanced computer networks allow integration of priority allocation functions of traffic into the end stations (computers) and into LAN and WAN communication equipment (primarily into switches and routers). Figure 3 shows a simple computer network, with several network stations (computers), and several communication devices that connect them.

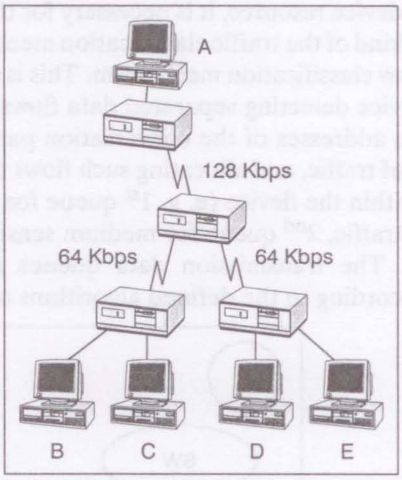

Figure 3 - Example of simple WAN network

If computer A wants to establish a connection of minimal throughput of 48 Kbit/s with computer B, and at the same time computer C wants to establish an extremely important communication connection of throughput 32 Kbit/s with computer D, it is necessary to install the priority function, since it is not possible to meet both requirements for communication at the same time. In this case it would be necessary for the computer C to require from the communication devices the allocation of priority and the quality of service, thus ensuring the priority in the transmission of this communication, and the communication between computers A and B would obtain the remaining part of the communication network throughput capacity, if the communication can support it.

This simple example illustrates the need for support to the allocation of priorities at end computers as well, since their requirements would not be satisfied if they did not know how to transmit their requirements to the communication equipment, that is, if the network did not distinguish between computers B and C, and their different needs for communication.

5. BASICS OF IMPROVING MULTIMEDIA DATA TRANSMISSION IN TCP/IP COMPUTER NETWORKS

The frequently used or recommended solutions to improve the quality of multimedia data transmission via TCP/IP computer networks are classified into four main categories:

5.1. Solution of the big product of delay and throughput capacity

The majority solutions from this area have resulted from the need to solve the problems of the older TCP/IP implementations. In older implementations, namely, there is a marked problem of solving the big product of delay and the throughput capacity due to the small size of the *"congestion window"* parameter, slow beginning of the *"sliding window"* algorithm, and therefore of the slow transmission recovery from network congestions. Also, the total transmission scalability depending on the number of simultaneous data flows is one of the major areas of improving the TCP/IP algorithm. Figure 4.4 shows the comparison of the scalability of the main TCP protocol implementations.

Some proposed changes include higher initial value of the TCP window, change of the "sliding window" algorithm, and introduction of the *selective acknowledgement* extension (SACK) which enables proper adaptation of the TCP network traffic source in case of loss of several network frames within one RTT period. Since the increase of the "congestion window" parameter is based on the increase of the number of frames, rather than value of the frame length in bytes, it is recommended that previously the maximum allowed frame size on individual network segment or connection is found.

Using the maximum allowed frame size affects the delay, but also maximally uses the throughput capacity of the network by reducing the influence of the frame overhead.

TCP algorithms of the flow and congestion control represent the difficulties for the multimedia data transmission due to the basic problem that the connection constantly increases the load on the network until congestion is reached, which is followed by sudden decrease of the load, with the process being iteratively repeated.

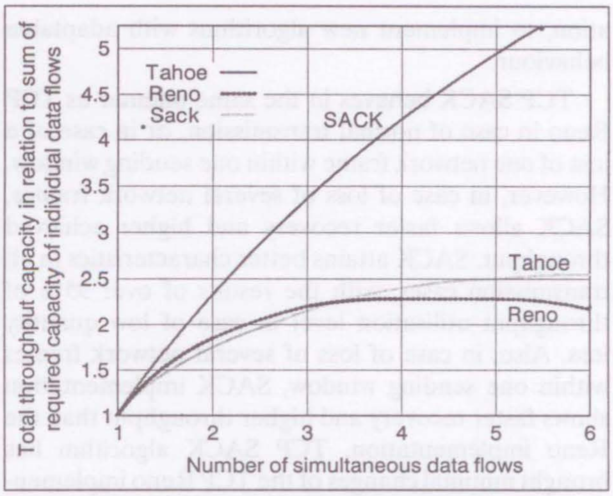

Figure 4 - Scalability of different TCP implementations depending on the number of simultaneous data flows

An example of load variation (throughput capacity) and the network congestion can be presented by two data flows that use the same transmission algorithm (e. g. TCP Reno), with the same network delay and beginning of transmission initiation. In this case both data flows will increase the load until reaching a half of the available throughput capacity, and the network will then reach the stable condition. In case that one of the data flows has greater delay than the other, the data flow with the lesser delay will increase its transmission window size, and therefore also the network load, significantly faster than the data flow with greater delay, and will therefore occupy also greater part of the network throughput capacity. In this way, the throughput capacity of the data flow with greater delay will be disturbed. Also, since in such a situation the total throughput capacity of the network is used, but the data flow with greater delay has not enough resources for normal transmission, both TCP Reno data flows will continue to fight for the available throughput, continuously bringing the network into the state of congestion.

Such behaviour of the TCP Reno implementation shows substantial lack of tolerance towards the connections with greater delays in the network. The effect of such behaviour on the multimedia communications can be extremely negative and may result in very low quality of video or audio signals in the network that has enough capacity to support these (regarding throughput capacity, delay or delay variation). Therefore, it is necessary to find better algorithms for delay identification and control within the TCP protocol implementation. Since the protocol operation frame allows introduction of various behaviour algorithms in the state of congestion, it is possible, in order to increase the efficiency and intelligence of the TCP operation, to implement new algorithms with adaptable behaviour.

TCP SACK behaves in the same manner as TCP Reno in case of normal transmission, or in case of a loss of one network frame within one sending window. However, in case of loss of several network frames, SACK allows faster recovery and higher achieved throughput. SACK attains better characteristics in all transmission cases, with the results of over 95% of throughput utilisation level in case of low quantity loss. Also, in case of loss of several network frames within one sending window, SACK implementation shows faster recovery and higher throughput than the Reno implementation. TCP SACK algorithm has brought minimal changes of the TCP Reno implementation. The key improvement refers to the facilitation of selective frame reception acknowledgements, and the setting of variable, which follows the number of frames remaining in the transit through the network. This provides smooth operation of the algorithm even in case several network frames have been lost within one data sending window.

The dependence of measuring the round trip time of network frames (RTT), as the main parameter for increasing the throughput capacity of sending the data flow in Reno and SACK implementations, is responsible for preferring the data flows with lower delay, i. e. RTT time.

In TCP Tahoe algorithm, special improvements refer to the detection mechanism, i. e. estimate of the round trip time of the network frames through the network. Also, *fast-retransmission* algorithm built into the Tahoe implementation was later modified in more advanced implementations. If this function was disabled, TCP Tahoe transmission algorithm would need substantial time in order to detect the loss of a network frame and to retransmit it to the network, which means that the quality of transmission in case of a single lost network frame would be disturbed. When the loss of frame is detected, the transmission continues with reactivation of the *slow-start* algorithm, which means that again there would be gradual increase in the sending throughput capacity.

5.2. Reduction of transmission errors

The only generally accepted solution to reduce errors in transmission is the use of FEC protocol (*Forward Error Correction*). However, FEC protocol cannot correct all the errors in transmission, since it requests additional throughput capacity and processing on the network nodes, which increases the delay and delay variation in the transmission. Further development of TCP SACK protocol plans a new approach to this area, possibly through a special type of ICMP *(Internet Control Message Protocol)* messages on data transmission errors.

5.3. Earlier acknowledgement of reception

One of the options of improving the performances is earlier sending of acknowledgements regarding reception of transmitted network frames. A popular method of improving performances with this approach is BECN (*Backwards Explicit Congestion Notification*) protocol, which is a variant of the ECN *(Explicit Congestion Notification)* protocol which uses ICMP messages and TCP Proxy system which temporarily saves network frames, and sends reception acknowledgements regardless of the actual receiver station. This mechanism is also called *link splitting.*

Over the last several years, several different methods of preventing the network segments of reaching the congestion area have been developed. These methods included the usage of special protocols for the detection of actually available characteristics of individual network segments (throughput capacity, delay and delay variation) and data flow control from the very end stations or control of multimedia data flows on the network nodes (routers). Among the former protocols the most popular are PTP and RTP *(Real Time Protoco)*, whereas the latter group contains protocols such as RED *(Random Early Detection)* and WRED *(Weighted Random Early Detection)*. Since the former group of protocols requires changes in the TCP/IP protocols at end stations, and support on the network nodes through the entire computer network, the latter group of protocols is much simpler to implement in practical computer networks.

6. CONCLUSION

TCP/IP family of network protocols was not originally designed to transmit multimedia data (especially real-time multimedia data). The Internet (which is based on the mentioned protocols) is today the most widespread standard in the communication networks, with the tendency of becoming the universal telecommunication standard in the world. The work analyses the possibilities of improving the characteristics of the Internet protocol in the transmission of multimedia data. By using various TCP protocol implementations, in different transmission conditions defined by various delays and delay variations, of available capacity throughput, and various loads in the multimedia data transmission, the necessary improvements have been determined as well as dynamic adaptations of the transmission algorithms, which enables the building of dynamically adaptable multimedia communication systems.

ARIJANA MACURA, M. Sc.
E-mail: arijana.macura@fpz.hr
DRAŽEN KOVAČEVIĆ, D. Sc.
E-mail: drazen. kovacevic@fpz. hr
University of Zagreb
Faculty of Transport and Traffic Engineering
Vukelićeva 4, 10000 Zagreb, Republic of Croatia
IGOR FRANOLIĆ, B. Eng.
Bolfan 44, 42230 Ludbreg, Republic of Croatia

SAŽETAK

INTEGRACIJA MULTIMEDIJSKOG PROMETA U RAČUNALNU MREŽU

U posljednjih nekoliko godina došlo je do rapidnog povećanja multimedijskog prometa u računalnim komunikacijskim mrežama. Nove primjene komunikacijskih mreža postavile su nove, velike zahtjeve na mreže, zbog potrebe da se korisnicima osigura kvalitetna usluga prijenosa multimedijskih sadržaja. Kao odgovor na takve zahtjeve razvijeni su brojni mehanizmi koji osiguravaju uslugu dodjele prioriteta prometu u komunikacijskim mrežama, kojom se postiže mogućnost određivanja kvalitete usluge koja je neophodna za pojedine mrežne aplikacije.

KLJUČNE RIJEČI

multimedijski promet, komunikacijske mreže, TCP/IP protokol, TCP algoritmi prijenosa, tokovi podataka

LITERATURE

[1] **Lenardić, D.; Zovko - Cihlar, B; Grgić, M**: *"Analysis of Effects of Network Latency on Multimedia Data Transfer in Broadband Communications Networks"*, EURASIP 2003, Zagreb, Croatia, 2003.

[2] **Lenardić, D.; Zovko - Cihlar, B; Rimac - Drlje, S.**: *"Analysis of Network Buffering Effects on TCP/IP Protocol Behavior"*, ELMAR 2003, Zadar, Croatia, 2003.

[3] **Fall, K.; Floyd, S.**: *"Simulation-based Comparisons of Tahoe, Reno and SACK TCP"* - Computer Communications Review, July 1996

[4] http://www.google.hr

[5] http://www.carnet.hr