Data-Rate and Queuing Method Optimization for Internetworking Medical Applications

Radek Dolezel, Otto Dostal, Jiri Hosek, Karol Molnar, and Lukas Rucka

Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 118, 61200 Brno, Czech Republic
xdolez35@stud.feec.vutbr.cz, otto@ics.muni.cz, hosek@feec.vutbr.cz, molnar@feec.vutbr.cz, rucka.lukas@phd.feec.vutbr.cz

Abstract. In medical environment, there is a fundamental demand to transfer and store large volumes of image data generated by modern medical devices. Currently the majority of the medical facilities spread around the country have quite limited Internet access. The aim of our work, presented in this article, was to find an optimal solution to transfer large volumes of image date over low-capacity links with regarding to minimum response-times. First we statistically described the traffic generated by the corresponding medical equipment and then evaluated the behaviour of these mathematical models in the OPNET Modeler discrete event simulation environment. The simulation results and their interpretation represent the main contribution of the following text.

Keywords: Computer Rentgen, Computer Tomograph, medical image processing, OPNET Modeler, transmission capacity, WFQ.

1 Introduction

Usually in regional medical image data processing systems, large volumes of data received from all cooperating medical facilities are stored in one central node. The sources of data, called modalities, are obviously MRI (Magnetic Resonance Imaging), CT (Computer Tomograph), US (Ultra-Sound) or CR (Computer Rentgen / X-ray) devices.

The usage of optical networks can provide sufficient bandwidth capacity for medical facilities, [1], [2]. Difficulties occur in medical facilities which are connected by alternative technologies with limited data-rate. The aim of this article is to find a solution which can compound the demands of hospital workers on the volume of required data and maximum acceptable delay.

The main goal is to find an optimal relation between a channel capacity and delay of images transmitted by various types of acquisition devices (modalities). Preferential treatment of some selected traffic-flows can also significantly affect the response-time of the evaluated services. Preferential treatment has its reason because not all of modalities are used for acute cases, so these modalities can have a less resources, e.g. data-rate in comparison with those used instantly. The simplified scheme of medical data transfer architecture is shown in Fig. 1.

R. Szabó et al. (Eds.): AccessNets 2010, LNICST 63, pp. 141–152, 2011.

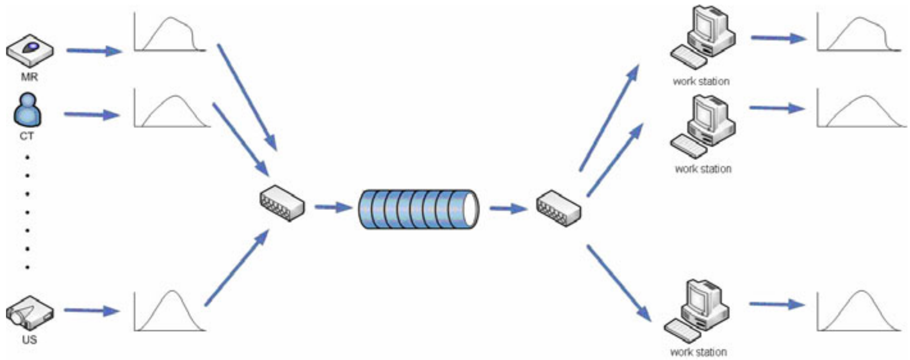

Fig. 1. Medical data transfer architecture – simplified scheme

2 Initial Premises and Statistical Analysis

The system in the scope of our evaluation uses TCP (Transmission Control Protocol) as a transport protocol, so the transfer time of image data can be affected by channel capacity, performance of the TCP transmitter and the receiver subsystem and by application functionalities.

We have experimentally verified that channel throughput is not limited by the size of socket-buffer neither in the transmitter nor in the receiver. So, there is no TCP window reduction caused by the lack of buffer capacity on the receiver or the transmitter side. We also assumed that the channel throughput was not influenced neither by the application behaviour such as the data storage and organization method. The parameters of a statistical model were specified based on the measurement and analysis of real traffic. The measurement was provided during traffic peaks, which is the time from morning 7 AM to 4 PM. As a source of investigated data the traffic from CT and CR modalities has been chosen. Three key traffic parameters have been identified which were required to model the traffic: inter-request time, size of transmitted data and number of repetitions. Since all of these parameters are random variables each of them were described by a corresponding probability distribution. The probability distributions of the corresponding traffic parameters have been tested by the Pearson's chi-square test with a significance level of 5%. The independence of the volume of transmitted data and the intervals between transmissions was verified by a contingency table.

To obtain the precise traffic-profile of the corresponding acquisition modality a precise long-term measurement has been carried out. We collected required data for one week, every day from 7am to 4pm. For this purpose the modalities connected with speed of at least of 100Mbps has been selected. The whole traffic from these modalities was captured using the tcpdump utility and subsequently analysed. The results of the analysis of the selected modalities are presented in the following chapter.

The CR was the firstly processed modality. First of all we analysed the inter-request time, e.g. whether during a given period of time the acquisition modality is transmitting data or not. From a practical point of view we found more useful to work with the periods between the establishments of the subsequent connections instead of the time between the end of the previous and the setup of the following connection. The reason is that the end of the connection depends on the capacity of the transmission links which is the parameter we want to optimize in our simulations.

Based on the analysis provided the probability distribution of the inter-request times between two subsequent TCP connections *DT* can be described by exponential distribution with parameter $\lambda = 1387.40 s^{-1}$. The amount of data transmitted *V* is a combination of two intervals with uniform distributions in ranges <11; 14> and <20; 95> MB. Values *DT* and *V* are independent.

The example of the slices of the secondly processed CT modality is shown in Fig. 2. During the analysis of the captured data a random number of TCP bursts have been identified during each relation. These bursts were usually represented by one to seven separate connections. Therefore, the connection time between bursts, the number of the TCP connections in the burst and the time between TCP connections in the burst were analyzed separately. A time interval of 150 seconds was set up as a time limit for TCP connection which is no longer considered to be a part of the investigated burst.

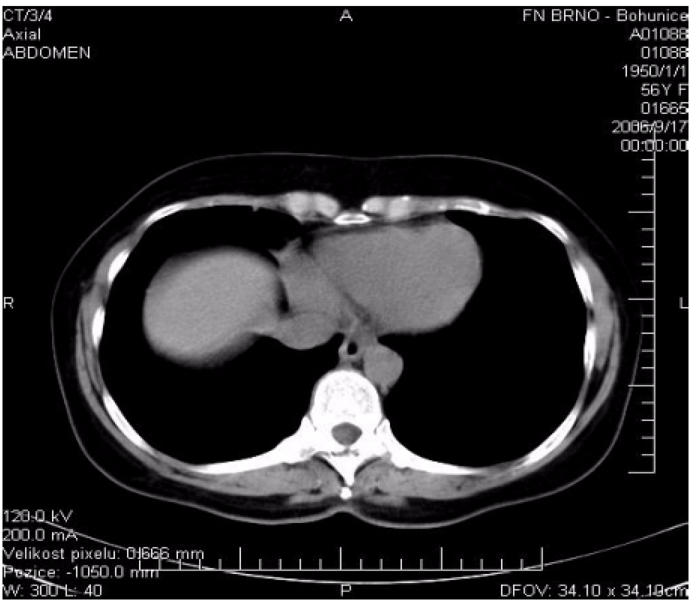

Fig. 2. One of the slices of the CT modality

The inter-request time between subsequent TCP connections *DT* has exponential probability distribution with parameter $\lambda = 837.63 s^{-1}$. The interval between TCP connections within one burst *DB* is in range from 8 to 150 seconds and has normal probability distribution with parameters $\mu = 57.87$ and $\sigma = 27.88$. Burst of the TCP connections contains from one to seven connections. The number of TCP connections in the burst NB has Poisson probability distribution with parameter $\lambda = 1.45$. The amount of data transmitted *V* in every TCP connection has an alternative probability distribution. The probability of transmitted data with 8.5MB in size is 0.25 and with 10.25MB in size is 0.75.

3 Simulation Results

Due to timing constraints in practical implementations we examined the impact of total link capacity and differentiated queue management on the response-time of the modalities. For this purpose a simulation model has been built in OPNET Modeler simulation environment [3]. The model consisted of four traffic sources modelling the CT modalities and other four traffic sources modelling the CR modalities. The topology of the simulation scenarios is in Fig. 3.

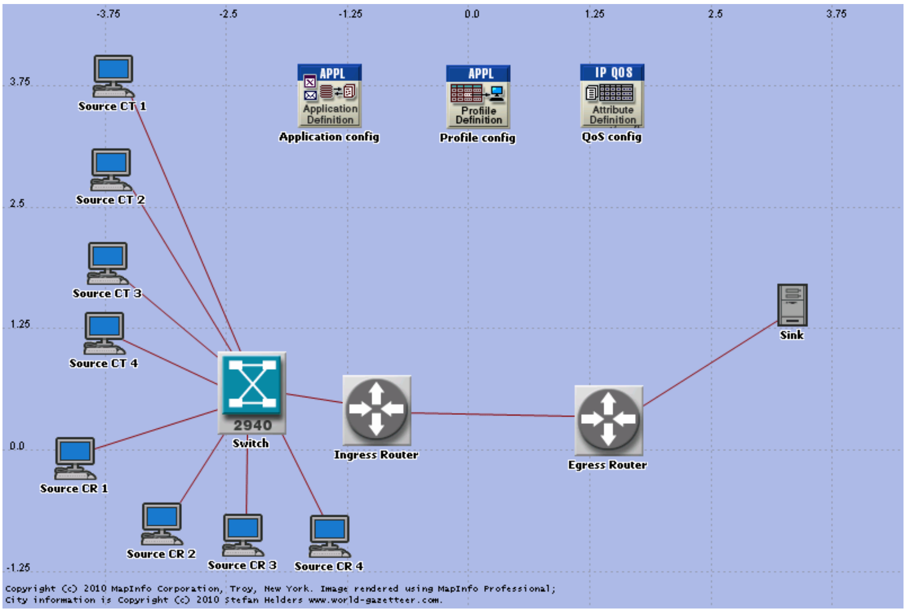

Fig. 3. Topology of the simulation scenario

During the simulations the application-level response-time has been evaluated. Because of a very close behavioural analogy, the FTP (File Transfer Protocol) protocol has been used to model both of the modalities. To simulate limited link capacities rate-limiting was applied on the common communication link. All the other communication links operated with full-speed 1Gbps. The inter-request time, file size and number of repetitions were configured according to the results obtained by statistical analysis of the captured traffic. In later simulation scenarios we also verified the influence of controlled queue management, namely the mechanism of WFQ (Weighted Fair Queuing) [4], [5]. The following figures show the most important simulation results.

Fig. 4 and Fig. 5 show the dependency of the response time (averaged for all four sources of the same modality) on the capacity of the rate-limited link. In the simulation the traffic was generated by all eight devices at the same time and the response times were averaged separately for each modality. For both modalities there is a significant increase in response times when the link-capacity is reduced through 10Mbps to 5Mbps. Based on simulation results, practically the capacity should not be dropped below 10Mbps, otherwise the quality of examined services will be markedly reduced.

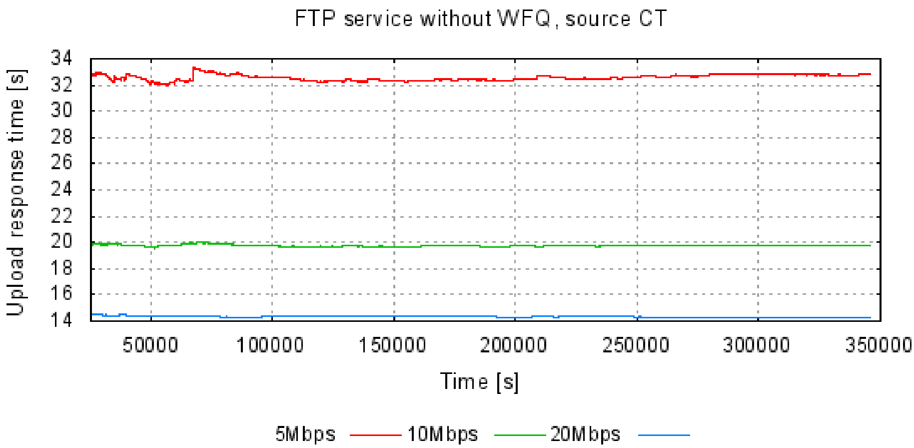

Fig. 4. Dependency of the average response-time of modality CT on the maximum link capacity

Fig. 11 evaluate the impact of WFQ on response-times. Two queues were used in the simulation the first one for one of the eight sources and the second for the remaining seven. It was necessary to distinguish between scenarios where the selected source is of modality CT (Fig. 6, Fig. 7 and Fig. 8) or CR (Fig 9, Fig. 10 and Figure 11). For both types of preferentially treated traffic-flows simulation scenarios with various maximum link-capacities were created. In addition, for each link-capacity four different bandwidth distribution models have been configured with different ratios between the bandwidth allocated to the first WFQ queue to the total bandwidth. More precisely the ratios of 20%, 30%, 50% and 80% have been used. For better comparison the corresponding graphs also contain the average response-time from the scenarios without WFQ queues.

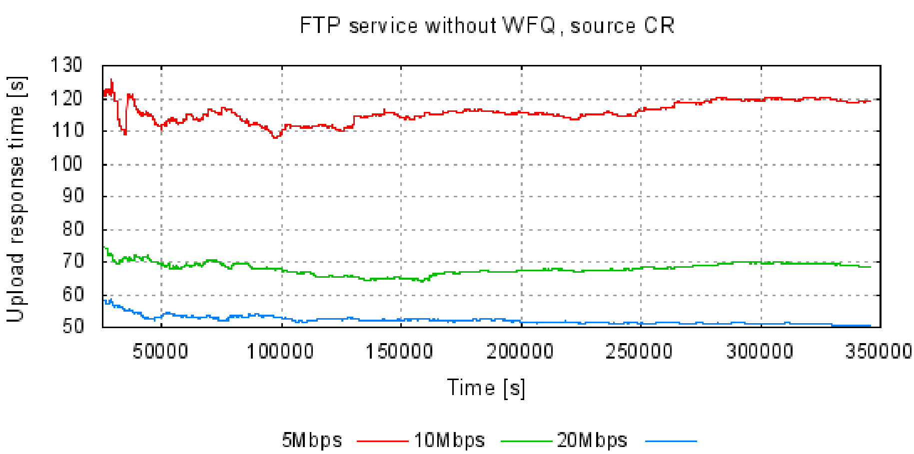

Fig. 5. Dependency of the average response-time of modality CR on the maximum link capacity

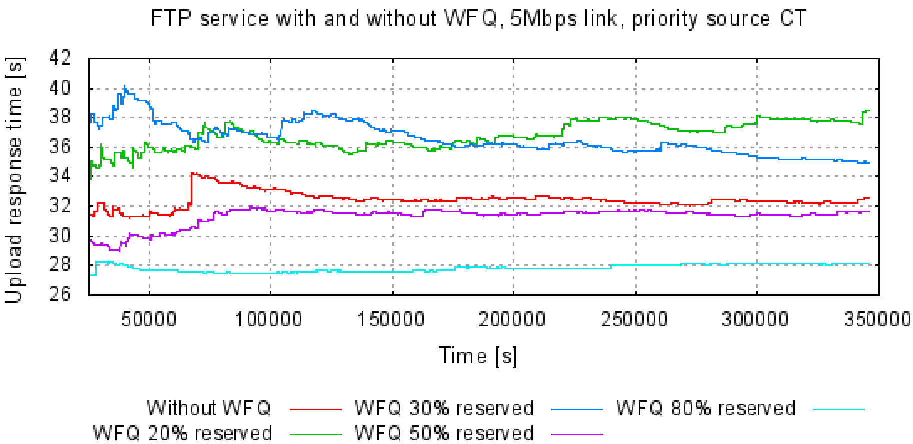

Fig. 6. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CT modality in the case of 5Mbps total link-capacity

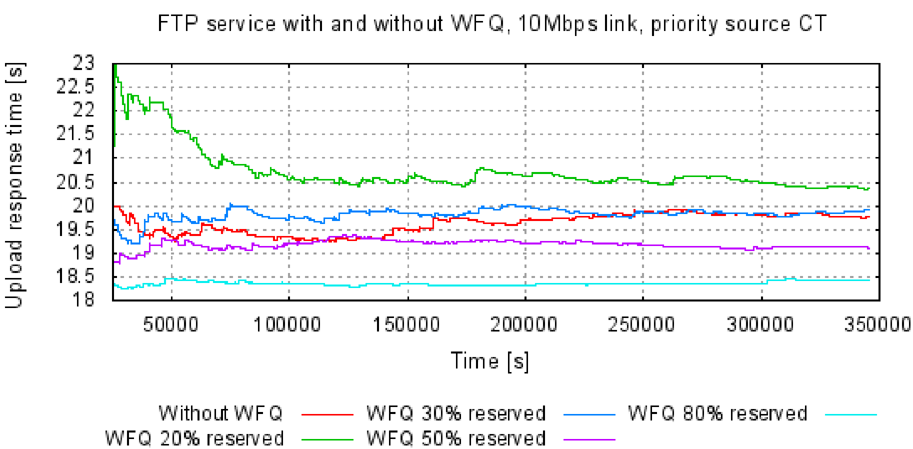

Fig. 7. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CT modality in the case of 10Mbps total link- capacity

Figures 6, 7 and 8 clearly show that WFQ improves the response-times for the CT modality. The influence of preferential treatment is more significant in the case of lower link-capacities, e.g. 5Mbps, see Fig. 6. In contrast, the impact of WFQ at a speed of 20Mbps with the given number of sources is practically negligible. Furthermore, the figures also show that the reduction of the response-times is significant only up to 30% of total bandwidth. Allocation of more bandwidth to one source brings no further improvements.

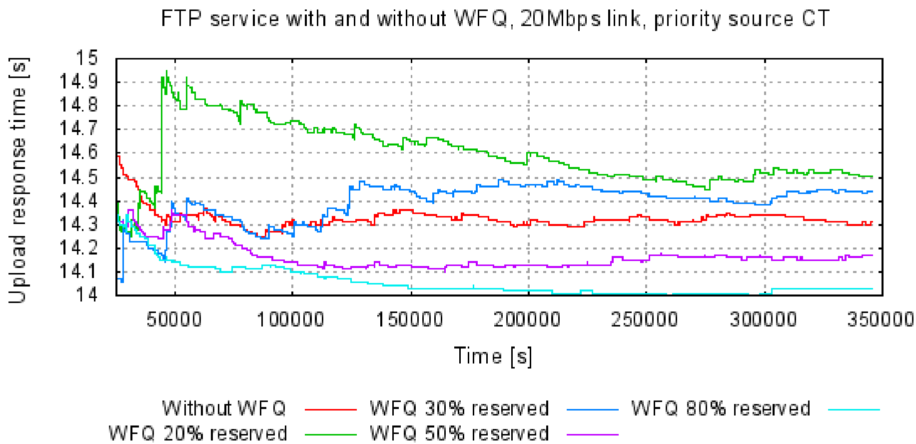

Fig. 8. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CT modality in the case of 20Mbps total link- capacity

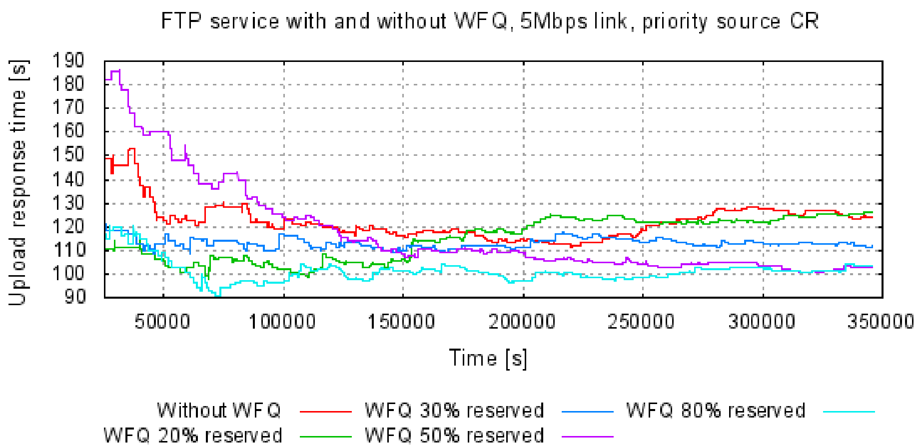

Fig. 9. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CR modality in the case of 5Mbps total link-capacity

Figures 9, 10 and 11 show the impact of WFQ on the response-times for a CR modality source. It is evident that for this modality WFQ does not bring any significant improvement nor at lower speeds. The reason of it is the bursty character of the CR modality. We can conclude that the efficiency of WFQ is highly dependent on the modality type and is not able to reduce the response-time under all circumstances.

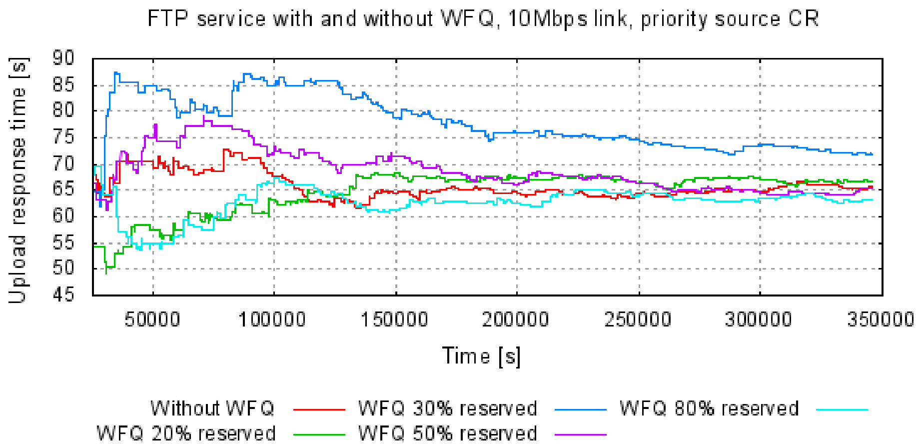

Fig. 10. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CR modality in the case of 10Mbps total link- capacity

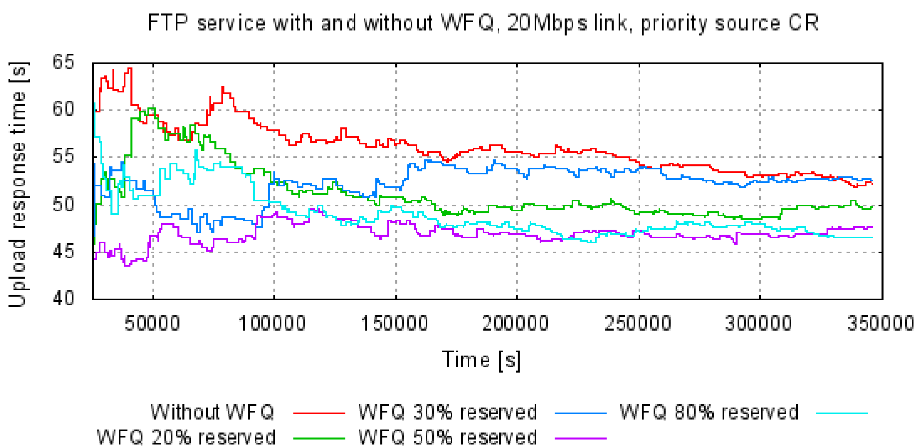

Fig. 11. Impact of the relative bandwidth distribution on the response-time of the preferentially treated CR modality in the case of 20Mbps total link- capacity

The following figures show a more detailed analysis of the impact of WFQ. Based on the earlier conclusions a bandwidth distribution model with 30% of resources allocated to the first queue (to the preferentially treated source) was used in the analysis. The results were divided based on the modality type of the preferentially treated source and the total link-capacity.

Fig. 12, 13 and 14 show the simulation results with preferentially treated CT modality in the case of 5Mbps, 10Mbps and 20Mbps total link-capacities respectively. There are five response-times included in each figure: 1) response-time of the preferentially treated source, 2) average response-time of the remaining three sources of the same modality, 3) average response-time of four sources of another modality, 4) average response time of the first modality without WFQ and 5) average response time of the second modality without WFQ.

From the results it is clear that the preferred traffic has shorter response-times than the others of the same modality, but this difference decreases by the increasing maximum link-capacity. Furthermore, at a speed of 30Mbps the preferential treatment of one source has a substantial negative impact on the average response time of the CR modality. This is due to the rarely generated but very large bursts of the CR modality, which in the case of bandwidth artificially limited to 70% of its original size cannot be transmitted as fast as in the case of standard best-effort treatment.

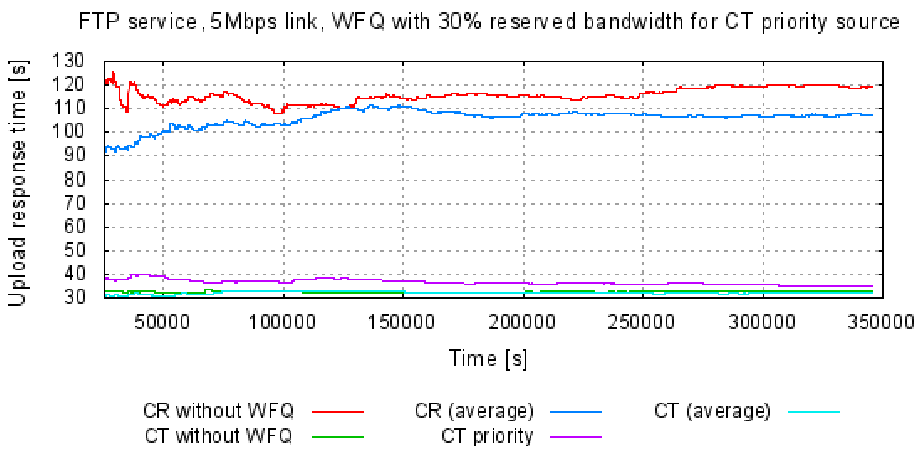

Fig. 12. Response-times when 30% of the 5Mbps link-capacity is reserved for one traffic source of CT modality

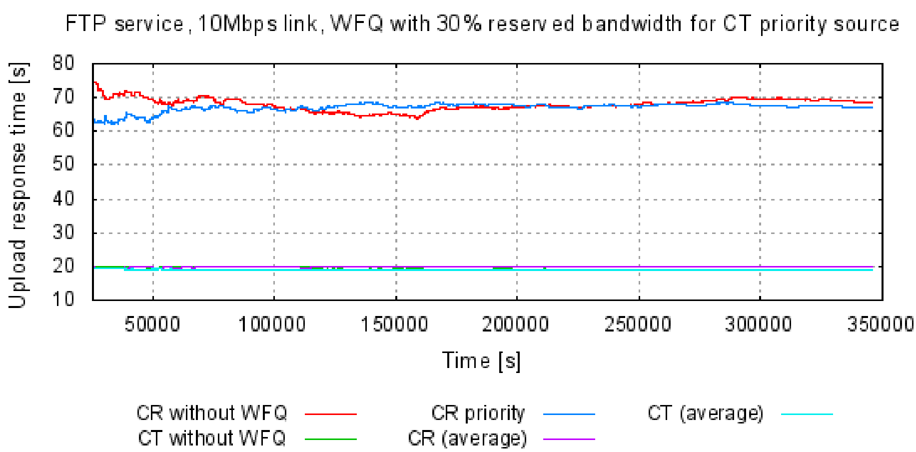

Fig. 13. Response-times when 30% of the 10Mbps link-capacity is reserved for one traffic source of CT modality

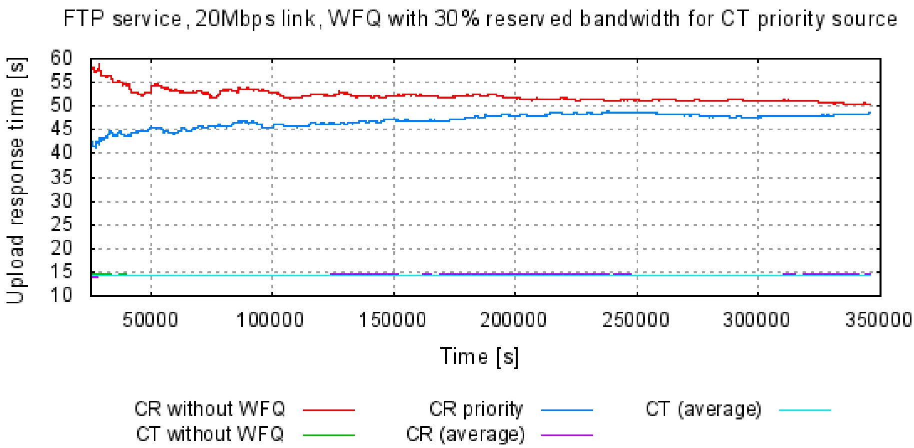

Fig. 14. Response-times when 30% of the 20Mbps link-capacity is reserved for one traffic source of CT modality

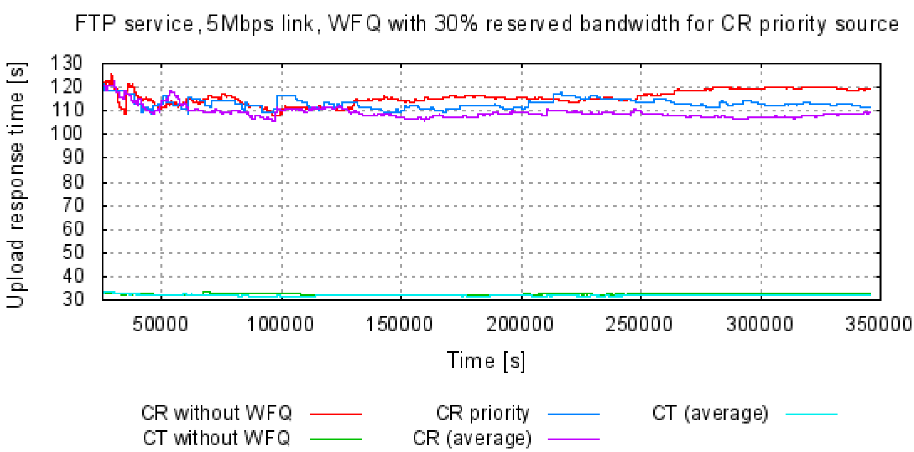

Fig. 15. Response-times when 30% of the 5Mbps link-capacity is reserved for one traffic source of CR modality

Figures 15, 16 and 17 show the simulation results with preferentially treated CR modality source in the case of 5Mbps, 10Mbps and 20Mbps total link-capacities respectively. As in the previous case, there are also five graphs in each figure: 1) response-time of the preferentially treated CR source, 2) average response-time of the remaining three CR sources, 3) average response-time of four CT modalities, 4) average response time without WFQ for the CT and 5) for the CR modalities. The results suggest that the impact of preferential treatment is evident only in a case of slow 5Mbps connection, see Fig. 15. In other situations the sorting of very large data burst into a queue with limited capacity seems rather counterproductive. The results also show that the preferential treatment of a CR modality practically has no effect on the response-time of the CT modality.

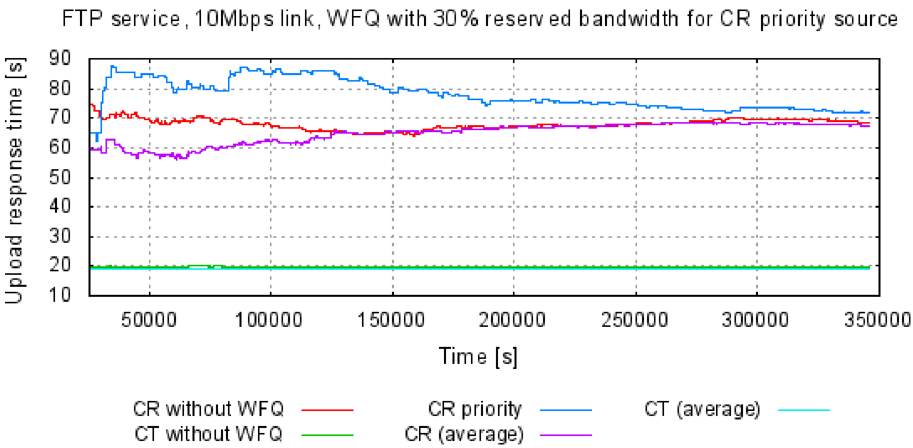

Fig. 16. Response-times when 30% of the 10Mbps link-capacity is reserved for one traffic source of CR modality

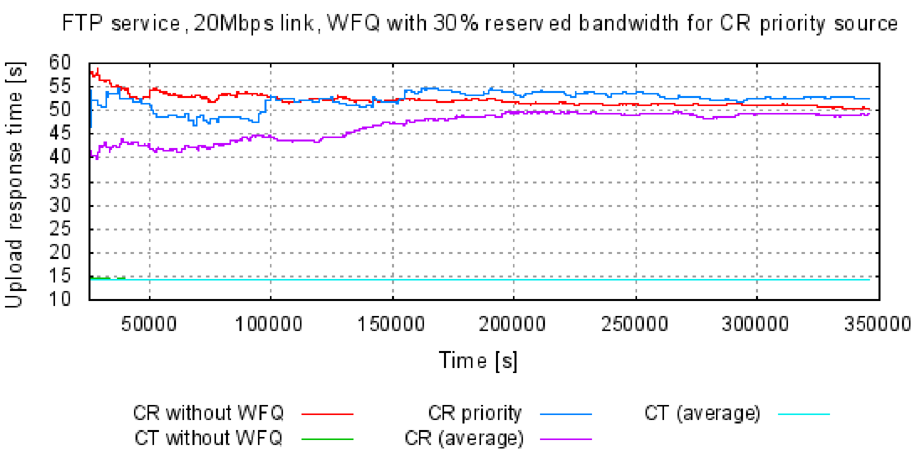

Fig. 17. Response-times when 30% of the 20Mbps link-capacity is reserved for one traffic source of CR modality

4 Conclusion

The aim of our work was to define a method how to estimate the link-capacity required by modern medical equipment communicating via data networks. Since hospital facilities are spread around the country usually they are interconnected trough commercial internet service providers and there is a natural pressure on minimization of the cost of these connections. On the other hand, long response-times can limit the practical usability of these equipments. Taking into account the previous constraints we suggested a two-step method, which firstly statistically describes the traffic generated by corresponding equipment and then evaluates the behaviour of these models in a discrete event simulation environment.

To verify the reliability of the method suggested we selected a mid size hospital facility with four CTs and four CRs. Next we derived their statistical model, based on data from long-term traffic-capturing, compared the response times calculated by OPNET Modeler with real values and found out that the simulation results substantially correspond to practical results.

To extend our analysis we also evaluated in the simulation environment the effect of quality of service support on response-times. During the analysis we confirmed that preferential treatment is significant only in the case of lower link-capacities, more exactly at 5Mbps maximum link-capacity for the selected combination of equipment. We also discovered that the efficiency of QoS support is highly dependent on the modality type and it is not able to reduce the response-time under all circumstances. This is caused by the bursty character of the modality. In addition, in some situations the bandwidth reservation appeared to be counterproductive as compared to standard best-effort treatment.

Acknowledgments. This paper has been supported by the Grant Agency of the Czech Republic (Grant No. GA102/09/1130) and the Ministry of Education of the Czech Republic (Project No. MSM0021630513).

References

1. Slavicek, K., Javornik, M., Dostal, O.: Technology backround of international collaboration on medicine multimedia knowledge base establishment. In: 2nd WSEAS International Conference on Computerr Engineering and Applications (CEA 2008), pp. 137–142. WSEAS Press, Acapulco (2008)
2. Slavicek, K., Novak, V.: Introduction of Alien Wavelength into Cesnet DWDM Backbone. In: 6th International Conference on Information, Communications and Signal Processing, pp. 61–66. IEEE Press, Singapore (2007)
3. Opnet Technologies: OPNET Modeler Product Documentation Release 15.0 (2009)
4. Park, K.I.: QoS in Packet Networks. Springer, New York (2004)
5. Růčka, L., Hosek, J., Molnar, K.: Advanced Modelling of DiffServ Technology. In: 32nd International Conference on Telecommunications and Signal Processing (TSP 2009), Asszisztencia Szervezo Kft., Budapest, Hungary, pp. 1–6 (2009)