Evaluation and Investigation of the Delay in VoIP Networks

Bakyt KYRBASHOV, Ivan BAROŇÁK, Matúš KOVÁČIK, Viktor JANATA

Inst. of Telecommunications, Slovak University of Technology, Faculty of Electrical Engineering and Information Technology, Ilkovicova 3, 812 19, Bratislava, Slovakia

kyrbashov@ktl.elf.stuba.sk, baronak@ktl.elf.stuba.sk, kovaci@ktl.elf.stuba.sk, janata@ktl.elf.stuba.sk

Abstract. *The paper is focused mainly on the delay problems, which considerably influence the final quality of connections in VoIP (Voice over IP) networks. The paper provides a detailed exploration of the nature and mechanisms of the delay. The main purpose of the investigation was an attempt to formulate a mathematical model of delay in the VoIP network and its subsequent analysis by laboratory data.*

Keywords

Delay, VoIP, jitter, IxChariot.

1. Introduction

Technologies dealing with the possibility of voice calls transmission using Internet Protocol (IP) data networks are summarily referred to as VoIP [1], [2], [3], [10]. Using data packet-switched network unlike the circuit switched network is a fundamental distinction from traditional telecommunication networks. Data packet-switched network in the most cases partially or completely does not guarantee the quality of services. Preference of IP networks as transport medium has brought that technology tremendous opportunities for simple and flexible extension in the world of computer networks. On the other hand, it is a source of large number of specific problems which VoIP technology carries.

The main issue in most IP networks is the inability to guarantee a certain quality parameters of transmission, especially network delay problems. The network delay plays a key function in the IP telephony, which may mostly influence the final quality of communication.

It is necessary to note that the new version of IPv6 takes place considerably advancing in Quality of Service (QoS) field [4], [7]. Likewise, several major manufacturers began to produce the active components of IP networks already with implemented support for IP telephony. Such devices facilitate to identify and prioritize handling of VoIP data.

Even though, the guarantee of VoIP connections quality is the basis for successful deployment and expansion of VoIP technology, obviously, the final quality of the connection is not only affected by the delay in the network, but also by many other parameters such as selection of a suitable compression algorithm, jitter, echo or packet sequence errors.

2. Delay Model of Voice Traffic in VoIP Networks

The delay is one of the very important factors that significantly affect the final quality of VoIP connections. In IP network we distinguish several types of delays that vary from their source, mechanism of creation and other features. Each component of the delay influences the resulting delay in a voice packet differently.

Delay components:

- $T_{\chi CDIP}$ – codec delay in IP environment,
- $T_{\pi PD}$ – packetization delay, which originates in voice data packetization,
- $T_{\beta QD}$ – queuing delay, which originate due to output queuing,
- $T_{\delta ser}$ – serialization delay,
- $T_{\omega PROP}$ – propagation delay,
- $T_{\Delta DJ}$ – de-jitter delay,
- $T_{\varepsilon PRi}$ – processing delay,
- $T_{\chi\upsilon DCD}$ – depacketization delay,
- $T_{CDa\lg}$ – algorithmic codec delay.

2.1. Codec Delay in IP Environment

Codec delay depends on the selection of codec and performance of the Digital Signal Processor (DSP). This is the time necessary for the DSP to compress PCM (Pulse Code Modulation) samples block. Each encryption algorithm specifies exactly how big block of PCM samples is coded together. This causes a delay, known as compression codec delay. This type of delay depends on the size of the block that is coded and it is exactly defined by the selected codec. The second partial delay, which occurs in coding process, is called algorithmic delay. Algorithmic delay depends on the functionality of coding algorithms, where some algorithms are required to encode the N-sample to know the N+1, N+2, ... samples. According to ITU-T Rec. G.114 there are different types of codec delays which depend on the applied environment [12]:

$$T_{\chi CDIP} = (N+1)T_{\chi CDa\lg} + T_{\chi CDcomp} \qquad (1)$$

where:

$T_{\chi CDIP}$ – delay caused by coder processing in IP environment,

$T_{\chi CDa\lg}$ – algorithmic codec delay,

$T_{\chi CDcomp}$ – compression codec delay,

N – number of frames per packet.

2.2. Packetization Delay

Packetization delay arises in the process of encapsulation of data block into packets that are transmitted over the network. Since the IP network is operating with variable packet size, it is possible to define what size of packets we want to generate in the system. And due to that we can affect how much data blocks from the codec have to be transferred together in a single IP packet.

Certainly, if we pass several blocks at once, it is necessary to hold them in the buffer until the time when the last block of data for particular packet are generated by the coder. Only then the data can be encapsulated and sent to the network. Packetization delay is set up in multiple packetization periods in which the given codec is operated.

In other words, it is an identification how many data blocks will be transferred in one packet.

The method of calculating the delay is given by (2)

$$T_{\pi PD} = \frac{8 \cdot P_S}{C_{BW}} \quad [\text{ms}] \qquad (2)$$

where:

$T_{\pi PD}$ – packetization delay [ms],

P_S – payload size [B],

C_{BW} – codec bandwidth [kbit/s].

2.3. Processing Delay

The processing delay in routers is the time necessary for a router to take packets from input interface to output buffer.

The delay of data processing in routers depends on various factors such as:

- central processor unit (CPU) speed,
- CPU utilization,
- IP switching mode,
- router architecture,
- configuration of input and output interfaces.

2.4. Bandwidth Variation

Each protocol of each transport technology used in IP networks involves an additive header that covers protocol information. These data are required by specific protocol or technology. The amount of information that is transferred as the protocol header depends on the type of the used technology. The type of technology used for transmission of packet can be different for each part of the final path. In these cases, when the change of technology along the path is needed, some headers are removed and replaced by new information needed for transmission in the new environment. The transport of additive information of course means the need of additive bandwidth. Amount of this bandwidth addition can be calculated. From the information about the amount of needed bandwidth for codec and the sampling interval of this codec, the amount of data part of IP packet can be calculated. This can be done by formula (3):

$$P_s = \frac{C_{BW}.T_s}{8} \quad [\text{B}] \qquad (3)$$

where:

P_s – payload size [B],

T_s – sampling period [ms],

$C_{BW.}$ – codec bandwidth [kbit/s].

After this calculation the amount of information allocated by the header can be added. Then the real needed bandwidth for this flow can be evaluated, as shown in formulae (4) and (5).

$$S_{BW} = \frac{H_L \cdot 8}{T_s} + C_{BW} \quad [\text{kbit/s}] \qquad (4)$$

where:

S_{BW} – total bandwidth [kbit/s],

H_L – header length [B],

T_s – voice sampling interval [ms],

$$S_{BW} = \frac{H_L \cdot C_{BW}}{P_s} + C_{BW} \quad [\text{kbit/s}]. \qquad (5)$$

2.5. Serialization Delay

Serialization delay is the component of the final delay, which depends on the transmission rate.

The operation of packet sending takes some amount of time. This time depends on the transmission speed and on the packet size. This time has a negative effect on final transfer delay. For evaluation of this time the formula (6) can be used.

$$T_{\delta ser} = \frac{F_s}{L_s} \quad [\text{ms}] \tag{6}$$

where:

$T_{\delta ser}$ – serialization delay [ms],

F_s – frame size is the total size of packet [B],

L_s – line speed [kbit/s],

and

$$F_s = H_L + P_s \quad [\text{B}] \tag{7}$$

where:

P_S – payload size, [B],

H_L – header length, [B].

2.6. Propagation Delay

This type of delay is related to physical transfer of signals in environments and materials. Propagation delay depends of course on the used technology and also on the distance along which the signal is transferred. For example in the case of distance evaluated in tens of kilometres the impact of this delay is negligible. The impact of this type of delay is significant in the case of transport networks (e.g. transoceanic or transcontinental paths, etc.), where long distance lines are used. In this type of networks the optic fibre is mainly used as the medium. The transmission of light can be described as transmission of electromagnetic wave in environment. The transmission rate depends on refractive index of this environment and can be calculated by the following formula:

$$v = \frac{c}{\eta} = 2{,}0 \cdot 10^8 \quad [\text{m.s}^{-1}] \tag{8}$$

where:

$c = 3 \cdot 10^8$ [m.s^{-1}] – speed of light in vacuum,

$\eta = 1.5$ – refractive index for silicium glass with wave length $\lambda = 1{,}33\ \mu m$,

v – speed of light spreading in optic fibre [m.s^{-1}].

Then we can get the value of propagation by the following formula (9):

$$T_{\omega PROP} = \frac{L}{v} \quad [\text{s}] \tag{9}$$

where:

$T_{\omega PROP}$ – propagation delay [ms],

L – length [m].

The final value of the propagation delay can be then around 4.38 microseconds per 1 km. This value is irrelevant in comparison with other values of delay on short distance transmissions. Only in the case of transcontinental transmission the value of propagation delay can be calculated in tens of milliseconds.

2.7. De-Jitter Delay

This kind of delay is in relation with jitter or delay variability. Sometimes it is necessary to eliminate impact of the delay by using a de-jitter buffer. This buffer can be considered as a memory. Most important is that size of this memory it is limited by delay variation and on the other hand, this type of jitter forms an additional part (dynamic) of delay. Then there is a need to optimize the size (length) of the de-jitter buffer. In a typical case the size of de-jitter delay is selected as 1.5-multiple of the sum of all variable factors of the delay. In the most cases if this value is set as static, values of de-jitter buffer size are somewhere in the range of 30 – 50 ms. If this value is set dynamically, it can be in the range of 100 – 150 ms. A more precise value can be set, but only after a deep analysis of the situation in the network. In our model of delay the value of de-jitter delay represents an additive factor of the final delay.

2.8. Depacketization Delay

Depacketization delay is similar to the delay of packetization. In case when one packet involves more data blocks, it is important to hold all blocks in buffer until the last of them is generated. Then the packet can be sent along the network. On the receiver side there is a similar situation. The receiver gets more blocks in one packet from the input. The first of these blocks can be decompressed immediately, but the other blocks have to wait in the buffer until they can be decompressed. It means that the first of these blocks gets zero delay and the last of them gets the value of packetization delay. These values are opposite to the values of packetization delay. In a real situation each block gets along the transmission path only one value of packetization delay.

2.9. Decompression Delay

The delay of decompression as well as the delay of compression is dependent on the choice of compression algorithm [3]. Average value of delay in decoder can be 10% of compression delay, in the most cases. But it is dependent on the computing power of the decoder and mainly on the number of data blocks in one packet. Mathematical representation of this type of delay can be described as:

$$T_{\chi\upsilon DCD} = 0{,}1 \cdot N \cdot T_{\chi CD} \quad [\text{ms}] \tag{10}$$

where:

$T_{\chi\upsilon DCD}$ – decodec delay [ms],

N – number of voice data blocks in one packet ,

$T_{\chi CD}$ – codec delay [ms].

2.10. Routing Delay

In network routers [17] multiple data and voice flows meet each other and they can be routed in different directions. It can cause that multiple voice and data flows can be routed through one output link. In this situation it is necessary to decide which packets will be prioritized and sent to the output link. By choosing the way of packet processing it is possible to significantly affect the overall delay. It is necessary for voice packets to be processed with priority over other data traffic. Voice packets can be labeled by mechanisms of differentiated services or integrated services [7], [11]. Processing in router is in most cases controlled by methods of PQ (Priority Queuing), or PQ/WFQ (Priority Queuing/Weighted Fair Queuing) [17] or by other queuing method.

2.11. Characterization of Overall Router Delay

Router traffic with PQ optimization is based on priority processing of packets in the primary voice PQ front [21]. In the case of ignoring the serialization delay of data packets with lower priority than voice packets, for traffic load modeling and router delay it is sufficient to observe the delay only in the priority queue. The method of handling the requirements of the priority queue corresponds to M/D/1/k queuing model, where k is the size of the buffer cache [6], [7]. For the purpose of formulating analytic model of a delay in the router, we can ignore buffer size and consider a system with sufficient buffer size without losses of priority packets. Presuming this fact we can replace the M/D/1/k model with M/D/1/∞ model. Then we can formulate analytical expression of probability that the buffer in router will be full. Consequently from this it is possible to formulate analytic model of packet delay in routers [22]. For M/D/1/∞ model, transition from one state can be only to the nearest state. This model corresponds to receiving or sending one packet.

Model can be applied with the following presumptions:

- *intervals between incoming individual requests correspond to Poisson probability distribution.* If we assume that only voice traffic from M sources is incoming to router model and every source has Poisson probability distribution [20], then their integration together will have Poisson probability distribution too,
- *$\lambda(t)$ – the probability density of service requirements is constant.* If we consider delay model with M incoming voice flows, that do not originate or vanish in time, then this presumption is fulfilled,
- *processing is handled by FIFO rule* (First In, First Out), standard router processing method in PQ queue corresponds to FIFO rule,
- *processing time is a constant parameter*, if we assume that M incoming voice flows will use the same compression codec and that they will generate packets with the same sizes, time of sending each packet to output link will be constant, therefore processing time will be constant too.

With these presumptions we can calculate system load from equation (11), i.e. output link load by voice packets.

$$\rho = \frac{\lambda}{\mu} \tag{11}$$

where:

λ – intensity of incoming requests $[s^{-1}]$,

μ – service rate $[s^{-1}]$,

ρ – system load.

For system stability, system load has to be: $0 \le \rho < 1$.

If we assume one traffic source with constant bit rate C_{BW} and constant packet data part size P_S, we can express intensity of the incoming requests as:

$$\lambda = \frac{C_{BW}}{P_S} \quad [s^{-1}]. \tag{12}$$

The service rate equals to:

$$\mu = \frac{1}{T_{\delta ser} + T_S} \quad [s^{-1}] \tag{13}$$

where:

$T_{\delta ser}$ – serialization delay of output link [s],

T_S – packet service time in router [s].

Then the probability of k requests waiting in the system can be expressed as:

$$p_k = (1-\rho)\sum_{j=1}^{k} \frac{(-1)^{k-j}(j\cdot\rho)^{k-j-1}(j\cdot\rho+k-j)e^{j\cdot\rho}}{(k-j)!} \quad \text{for } k \ge 2,$$

$$p_k = (1-\rho)(e^{\rho}-1) \quad \text{for } k = 1,$$

$$p_k = (1-\rho) \quad \text{for } k = 0. \tag{14}$$

The mean waiting time of request in the system can be expressed by (15):

$$T = \frac{1}{\mu} + \frac{\rho}{2(1-\rho)\mu} \quad [\text{s}] \tag{15}$$

where $1/\mu$ is service time of one request, and then mean number N of requests in the system can be expressed as:

$$N = T \cdot \lambda \qquad [-]. \tag{16}$$

Assuming that M voice flows are incoming to router and using Poisson probability distribution properties, where by integration of multiple voice flows with Poisson probability distribution together we get data flow also with Poisson probability distribution, whose λ parameter equals to the sum of λ_i parameters of individual voice flows, we can express the intensity of incoming requests as:

$$\lambda = \sum_{i=1}^{M} \frac{C_{BWi}}{P_{Si}} \qquad [s^{-1}]. \tag{17}$$

Assuming that all voice flows incoming to router use the same codec and all of them are transferred through the same transport environment, we can simplify (17) to:

$$\lambda = M \frac{C_{BW}}{P_S} \qquad [s^{-1}]. \tag{18}$$

If we know the transfer speed of an output link and the packet service time for particular router, we can put equations (5) and (6) to (13) and specify the service rate of the system as follows:

$$\mu = \frac{L_S}{P_S + H_L + L_S \cdot T_S} \qquad [s^{-1}] \tag{19}$$

where:

H_L – packet header size,

L_S – link rate,

T_S – packet service time in the router.

After substitution of equations (18), (19) to (11) we get the following expression for the system load:

$$\rho = \frac{M \cdot C_{BW}(P_S + H_L + L_S \cdot T_S)}{P_S \cdot L_S}. \tag{20}$$

By substitution of (19), (20) and (21) to (17) we get the expression of mean service time in the following form:

$$T = \frac{1}{2} \cdot \frac{P_S + H_L + L_S \cdot T_S}{L_S} \cdot \frac{2P_S \cdot L_S - C_{BW}.M(P_S + H_L + L_S \cdot T_S)}{P_S \cdot L_S - C_{BW}.M(P_S + H_L + L_S \cdot T_S)} \quad [\text{ms}]. \tag{21}$$

Similarly after substitution of (17), (18) and (19) to (14) we can express the probability of k requests in the system as follows:

$$p_k = \left(1 - M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{L_S \cdot P_S}\right) \cdot \sum_{j=1}^{k} \left[(-1)^{(k-j)} \left(j \cdot M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{P_S \cdot L_S} \right)^{(k-j-1)} \cdot \left(j \cdot M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{P_S \cdot L_S} + k - j \right) \cdot \frac{e^{j \cdot M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{L_S \cdot P_S}}}{(k-j)!} \right] \quad \text{for } k \geq 2,$$

$$p_k = \left(1 - M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{L_S \cdot P_S}\right) \cdot \left(e^{j \cdot M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{L_S \cdot P_S}} - 1 \right) \quad \text{for } k = 1,$$

$$p_k = \left(1 - M \cdot C_{BW} \cdot \frac{P_S + H_S + L_S \cdot T_S}{L_S \cdot P_S}\right) \quad \text{for } k = 0. \tag{22}$$

Then if we express service time from service rate equation (19), we have the probability of delay:

$$p_{T_k} = p_k \cdot \frac{P_S + H_L + L_S \cdot T_S}{L_S \cdot P_S} \quad \text{for k} = <0, \infty>. \tag{23}$$

Equation (21) which depicts an average packet delay in the router, together with equation (23), form mathematical model of the delay in the router. With this model we can calculate packet delay in the router for a chosen configuration of input parameters.

It is necessary to realize that this model of packet delay in the router includes serialization delay of the output link. In overall delay gained from this model, serialization delay of the output link is directly included in the calculation.

2.12. Consistent Mathematical Model of the Overall Delay in VoIP Network

Using the information from previous chapters, we can express the model of overall delay in the network as follows. Suppose the network structure illustrated in Fig. 1. There are two communicating endpoints connected through 4 routers with 5 different lines. The network carries voice and data communications and for simplification we assume that all voice streams are using the same compression algorithm with the same settings.

Endpoints are specified by used codec and the size of the data packet that is generated. Each line is specified by its bit rate, used transmission protocols, transmission technologies and their lengths. Routers in the network are specified by their processing delay and they are serving M_i concurrence connections. The receiver buffer is intended for jitter compensation, which is specified by its size.

Based on previous information we can express the total delay as follows:

$$T = (N+1)T_{\chi CDalg} + T_{\chi CDcomp} + T_{\pi PD} + \sum_{i=1}^{5} T_{\delta ser1} + \sum_{i=1}^{5} T_{\omega PROP1} + \sum_{i=2}^{5} T_{\beta QD2} + \sum_{i=2}^{5} T_{\varepsilon PR2} + T_{\Delta DJ} + T_{\pi\mu DPD} + T_{\chi\upsilon DCD} \quad [\text{ms}] \tag{24}$$

Then, after substitution we get a mathematical model of the delay in the network:

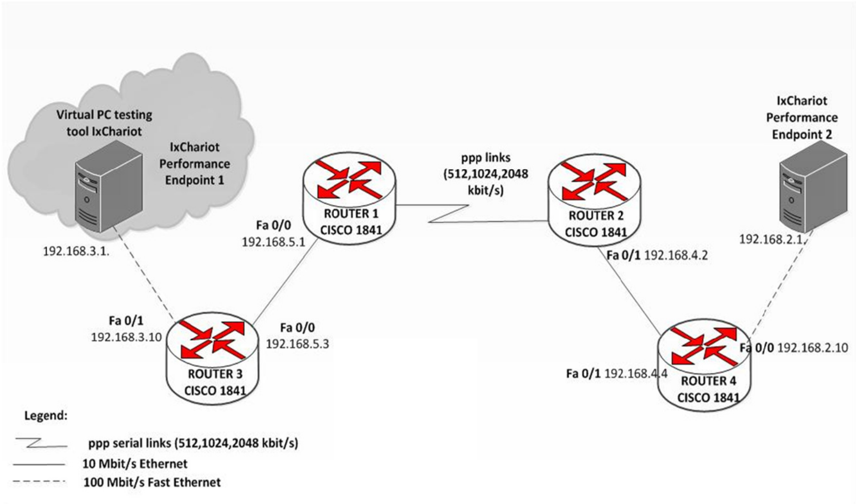

Fig. 1. Topology of experimental measurement system.

$$
\begin{aligned}
T &= (1+0{,}1\cdot N)(N+1)T_{\chi CDa\lg} + T_{\chi CDcomp} + \\
&+ \frac{8\cdot P_S}{C_{BW}} + \sum_{i=1}^{5}\frac{L_i}{\nu} + \sum_{i=2}^{5} T_{\varepsilon PRi} + T_{\Delta DJ} + T_{\pi\mu DPD} + \\
&+ \frac{1}{2}\sum_{i=2}^{5}\left[\frac{P_S + H_{Li} + L_{Si}\cdot T_{Si}}{L_{Si}}\cdot\right. \\
&\left.\cdot\frac{2P_S\cdot L_{Si} - C_{BW}\cdot M_i(P_S + H_{Li} + L_{Si}\cdot T_{Si})}{P_S\cdot L_{Si} - C_{BW}M_i(P_S + H_{Li} + L_{Si}\cdot T_{Si})}\right] \text{ [ms]}
\end{aligned}
\tag{25}
$$

where:

N – number of voice blocks in packet,

T_{CDalg} – algorithmic codec delay [ms],

T_{CDcomp} – compression codec delay [ms],

P_S – payload size [B],

C_{BW} – codec bandwidth [kbit/s],

H_{Li} – header length [B],

L_{Si} – speed of i^{th} line [kbit/s],

L_i – length of i^{th} line [m],

ν – speed of light in vacuum [m.s^{-1}],

$T_{\Delta DJ}$ – de-jitter delay [ms],

M_i – number of simultaneous calls directed to a router i^{th} output link,

$T_{\varepsilon PRi}$ – processing delay [ms],

$T_{\pi\mu DPD}$ – decodec delay [ms].

The formula (25) is the mathematical model of the delay in IP network of defined structure. The model can be generalized to any known network topology. Then if we know the traffic in the network we can calculate the average delay of voice traffic.

3. Emulation and Measurement

As an optimal way of realizing stated measurements this method was chosen: on a known network structure the sufficient number of voice connections is generated and every variable is observed. Because this experiment was realized for the purpose of verification of the formulated mathematical model of delay, which includes coding delay, processing delay, packetization delay and decoding delay, it was necessary to find a way to consider the whole chain of delays in the experiment as they appear in real environment, not just measure the delay of the network itself. Therefore it was necessary to determine a suitable way of observing VoIP voice canals, a way of exact monitoring of end-to-end delay of each individual voice connection. For this reason IxChariot software by Ixia was chosen [13].

This software allows you to generate an exactly defined traffic and also offers a detailed analysis of this traffic. It supports creation of multiple VoIP voice connections with specifically defined variables. IxChariot environment allows you to observe individual variables of each connection, such as delay, packet loss, jitter, MOS and R-factor. This product works on this principle: one central console can control more endpoints that generate real traffic in the network. During the test time, endpoints collect information about an actual state of transmission. After the whole measurement endpoints send the information to the central console, which provides post process interpretation. In the experiment we have generated voice flows using G.711a codec and data flows so that the transmission link load was at 80 %. Topology of the experimental measurement system is shown in Fig. 1. The PPP link rates between Router1 and Router2 were 512, 1024 or 2048 kbit/s and also fragmentation variables were changing for G.711a codec in relation to link rate and data part of G.711a codec.

For the G.711a codec were carried out tests for periods of framing values of 20, 40, 60 ms.

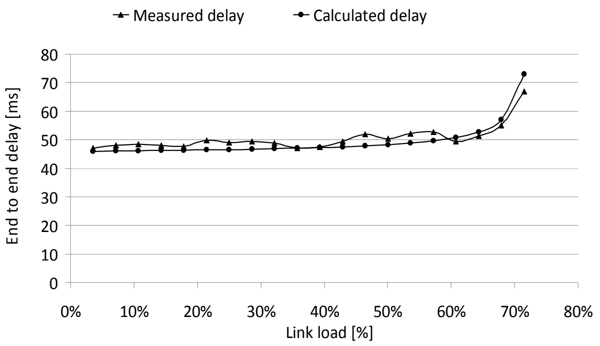

Fig. 2. Deviation between calculated and measured delay in relation to link load for G.711a codec.

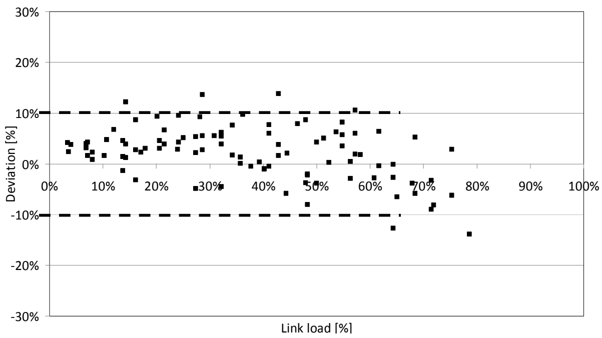

Fig. 3. Dependence calculated deviation in relation to link load for G.711a codec.

For each link rate setting tests were made for one, two, three, etc. concurrent voice connections. Maximum number of concurrent voice connections was set to PPP link rate, so that it would not be overloaded and no packets would be lost. For each link rate and number of connections multiple tests were made for different values of frame period. In one test, one type of codec voice flows were generated with the same frame setting for all voice connections and data flows. Each test took 2 minutes; monitored variables were recorded in one second intervals.

For error minimization each test was repeated three times with same the settings. From these 3 measured values arithmetic average value was calculated. Router1 and Router2 were set to strictly prioritize voice connections.

3.1. Evaluation of the Data

All the information gathered from the tests was summarized, statistically evaluated and then confronted with calculated values from the mathematical model of delay. The aim was to verify the calculated values and determine the working parameters of the interval in which this model provides reliable results. Fig. 2 illustrates individual measured and calculated values of delay for 2048 kbit/s link rate and G.711a codec.

In this configuration the number of concurrent connections increased from 1 to 22 for 60 ms frame setting.

After a deep inspection of the results, a more appropriate form of representation of results was chosen, in which the deviation between the calculated and the measured value is in relation with link load. This kind of expression of results was chosen because it displays the accuracy of values gained from the mathematical model of delay. Fig. 3 represents the percental deviation between calculated and measured values in relation to link load for all three link rates (512, 1024 and 2048 kbit/s).

The form of the relation to link load was chosen based on the presumption that with an increasing link load some kind of bursts in voice traffic will be more visible, but that was not included in the mathematical model.

4. Conclusion

In the formulated mathematical model an approximation of voice traffic was used. Generator of traffic had Poisson probability distribution. This does not exactly match the real properties of voice traffic, mainly its bursty character. Therefore the presumption was that with increasing link load the mathematical model will not give sufficiently accurate information.

Results gained from the mathematical model should differ from real traffic when the link load is more than 70 %. However, measurements have shown that in the most cases the formulated mathematical model gives results with ± 10 % accuracy up to 80 % link load. With increasing number of concurrent voice calls and decreasing link load, the accuracy of the gained results is better. The reason of this is that even if individual voice flows do not match the exact model of generator with Poisson probability distribution, their sum will converge to this model with increasing number of calls. Considering that in the most projected VoIP networks the number of concurrent connections is much higher, presumption is that stated model will provide sufficiently accurate values of average delay in the network.

Acknowledgements

This work is a part of research activities conducted at Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Institute of Telecommunications, within the scope of the projects ITMS 26240120029, “Support of Centre of Excellence for SMART technologies, systems and services II”, co-funded by the ERDF.

References

[1] KHASNABISH, B. *Implementing Voice over IP*. New York: Wiley, 2003.

[2] DAVIDSON, J., PETERS, J., BHATIA, M., KALIDINDI, S., MUKHERJEE, S. *Voice over IP Fundamentals.* 2nd Ed. Cisco Press, 2006.

[3] RAMESH, K.,ASADULLAH,S. *Cisco IP Telephony: Planning, Design, Implementation, Operation, and Optimization*. Cisco Press, 2005.

[4] CHROMÝ, E., KAVACKÝ, M. Asynchronous networks and Erlang formulas. *International Journal of Communication Networks and Information Security (IJCNIS),* 2010, vol. 2, no. 2, p. 85 – 89.

[5] JEREMIAH, F., THIMMA, V., BABU, G. *Modeling and Analysis of Telecommunications Networks*. New York: Wiley, 2004.

[6] BOLCH, G., GREINER, S., MEER, H., KISHOR S. *Queuing Networks and Markov Chains*. New York: Wiley, 2006.

[7] PARK, K. *QoS in Packet Networks*. New York: Springer Science + Business Media, 2005.

[8] BARONÁK, I., HALÁS, M., ORGON, M. Mathematical model of VoIP connection delay. In *Telecommunications, Networks and systems: International Conference*. Lisbon (Portugal), 2007, p. 95 – 99.

[9] GHANASSI, M. KABAL, P. Optimizing voice-over-IP speech quality using path diversity. In *IEEE 8th Workshop on Multimedia Signal Processing*. Victoria (USA), 2006, p. 155 – 160.

[10] KARAPANTAZIS, S., PAVLIDOU F. VoIP: A comprehensive survey on a promising technology. *Computer Networks*, 2009, vol. 53, no. 12, p. 2050 – 2090.

[11] VOZŇÁK, M., HROMEK, F. Optimization of VoIP service queues. In *Proceedings RTT2008, 9th International Conference*. Vyhne (Slovakia), 2008.

[12] ITU-T Recommendation G.114 (05/2003). *One Way Transmission Time*. Geneva: ITU, 2003.

[13] Ixia. *IxChariot*. [Online] available at: http://www.ixiacom.com/ products/ .

[14] BROSH, E., BASET, S., MISRA, V., RUBENSHTEIN D., SCHULZRINNE, H. The delay-friendliness of TCP. In *Joint International Conference on Measurement and Modeling of Computer Systems*. Annapolis (MD, USA), 2008.

[15] SILVA, A., VARELA, M., SILVA, E., ROSA M., RUBINO, L. Quality assessment of interactive voice application. *Computer Networks*, 2008, vol. 52, no. 6, p. 1179 – 1192.

[16] MEDVECKÝ M., BALOGH, T. Packet prioritization in networks with QoS support. *EE – Journal for Electrical Engineering and Information -Technology*, 2007, vol. 13, p. 208 – 211.

[17] NIKKEL, B. An introduction to investigating IPv6 networks. *Digital Investigation*, 2007, vol. 4, no. 2, p. 59 – 67.

[18] JOSEPH, V.,CHAPMAN B. *Deploying QoS for Cisco IP and next generation networks*. Burlington (MA, USA): Elsevier, 2009.

[19] VAJÍČEK, M., MEDVECKÝ, M. Evaluation of transport technologies for NGN transport networks. *EE – Journal for Electrical Engineering and Information – Technology*, 2009, vol. 15, p. 164 – 168.

[20] FISCHER, W., MEIER-HELLSTERN, K. The Markov-modulated Poisson process (MMPP) cook book. *Performance Evaluation*, 1992, vol. 18, p.149 – 171.

[21] FERRARI, T., PAU, G., RAFFAELLI, C. Measurement based analysis of delay in priority queuing. In *Proceeding of IEEE Global Telecommunication Conference Globecom 2001*. S.Antonio (TX, USA), 2001.

[22] FU, J., HAGSAND, O., KARLSSON, G. Queueing behavior and packet delays in network processor system. In *15th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, MASCOTS '07*. Istanbul (Turkey), 2007, p. 217 – 264.

About Authors...

Bakyt KYRBASHOV was born in Kyrgyz Republic in 1981. He graduated from Kyrgyz State Technical University in 2003. Since 2003, he was working at Kyrgyztelecom, an engineer at Development Department. In 2005, after his experience he has granted the scholarship of Slovak Government to study the PhD program in Institute of Telecommunications STU in Bratislava. In 2009 he submitted his PhD work in the field of delay in VoIP networks. Nowadays he works as a researcher in Institute of Telecommunications of FEI STU in Bratislava.

Ivan BAROŇÁK was born in Žilina, Slovakia in July 1955. He received the electronic engineering degree from Slovak Technical University Bratislava in 1980. Since 1981 he has been a lecturer at Institute of Telecommunications, STU Bratislava. Nowadays he works as a professor at Institute of Telecommunications of FEI STU Bratislava. Scientifically, professionally and pedagogically, he focuses on problems of digital switching systems, ATM, Telecommunication management (TMN), NGN, VoIP, QoS, problem of optimal modeling of private telecommunication networks and services.

Matúš KOVÁČIK was born in Trnava, Slovakia in November 1983. He achieved the electronic engineering degree at the Slovak Technical University in Bratislava in June 2008. Since April 2008 he has been a technician at the Institute of Telecommunications, STU Bratislava. Nowadays he is a researcher there and works on his PhD in the field of QoS in NGN.

Viktor JANATA was born in Bratislava, Slovakia, in 1982. He received his M.Sc. degree in Telecommunications from Faculty of Electrical Engineering, University of Žilina, Slovakia in 2006. Since March 2007 he is a Ph.D. student at the Institute of Telecommunications, Slovak Technical University in Bratislava, Slovakia. His areas of interests include IP networks, QoS parameters and modeling of telecommunications networks.