VoIP performance on Local Area Networks

Anton Kos, Aleš Vugrinec and Sašo Tomažič

Abstract—This paper investigates the possibilities of real-time voice transmission over local area networks networks that provide no guarantees of quality of service. The network of choice is 10 Mbit/s Ethernet and voice coder of choice is G.723.1. Both are widely used. At the beginning we define the requirements for real-time voice transmission and give the basic properties of G.723.1 coder. Delay is identified as the key parameter to real-time voice transmission. Then we identify the most common simplifications that are introduced on the way to theoretical results. We present results of a bimodal delay analysis that is tailored to the transmission of VoIP frames. We identify the upper bounds of network utilisation under which the delay is within acceptable range. Next we present the measurement results for different levels of network utilisation and compare them with simulation results that matched the measured network. Results of both are very similar and demonstrate that simulations are a good tool to predict VoIP behavior on the network. Despite the fact, that theoretical results were not directly comparable with measurements and simulations, they still give good guidelines about network operation under different conditions. All three methods thus complement each other.

Index Terms—VoIP, real-time voice transmission, VoIP simulation, VoIP measurements, VoIP performance Ethernet

1 INTRODUCTION

THE purpose of this paper is to investigate the possibility of voice transmission of sufficient quality over Ethernet networks with the help of theoretical analysis, simulations and measurements on a real system. As elements of the system we chose standards and technologies most used for this purpose. The Ethernet network considered is IEEE 802.3 10BaseT, and the voice coder ITU G.723.1. We chose 10BaseT for various reasons. During the implementation of simulations and measurements, this technology was dominant in terms of the number of connections. In addition, it also represents an extreme case in terms of delay, since more modern versions of Ethernet, such as 100BaseT and others, are better in this respect. We chose G.723.1 because this is the only voice coder available in all considered VoIP devices.

First we evaluate the capabilities of Ethernet networks and recognise their advantages and limitations. Here we pay most attention to analysis of delays under different network loads and different proportions of short frames to long frames, since this parameter is the key to real-time voice transmission. Next, we discuss the results of measurements on a real system, which we implemented in the laboratory, and compare them to simulations of the same measuring system. The analysis is not limited to general analysis of Ethernet frame delays, but we also provide an analysis of voice frame delays at different levels of network utilisation.

2 VOICE TRANSMISSION REQUIREMENTS

Real-time voice transmission is a challenging task for networks, since they must meet the following requirements, some of which are intertwined:

- sufficient available bit rate,
- sufficiently low delays,
- echo cancellation,
- bandwidth reservation, and
- low bit-error rate.

Some of these requirements can be defined with absolute values, while others depend on the coding methods used and the properties of the transmission network. For ***real-time transmission***, simultaneous compliance with the bit-rate, bandwidth reservation and particularly delay criteria is essential. Let us look in more detail at the delay criterion, which is the most important in terms of voice transmission over Ethernet networks[1].

Delay must be within certain limits, otherwise the quality of voice and interactivity of communication declines significantly. The upper limits of acceptable delay for voice transmission under recommendation ITU G.114 [9] are:

- delay of up to 150 ms is suitable for most user and voice applications,
- delay of between 150 and 400 ms is acceptable, provided that the network operator and the user are aware of the impact of increased delay on the quality of voice,
- delays above 400 ms are unacceptable for most user and voice applications.

3 BASIC PROPERTIES OF VOICE CODER

The G.723.1 voice coder is part of the wider set of standards H.323, which are intended for real-time data transmission of multimedia applications over local networks that provide no guarantees of QoS. Transmission of coded speech uses protocol stack RTP/UDP/IP, which adds a further 40 octets of overhead to a speech frame 24 octets long. For G.723.1, a typical value of delay due to speech processing is 67.5 ms. This consists of the time required to generate a speech signal belonging to one speech frame (30 ms), look-ahead in the

- *Anton Kos is with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. E-mail: anton.kos@fe.uni-lj.si.*
- *Aleš Vugrinec is with the Ministry of Defense, Slovenia. E-mail: ales.vugrinec@gov.si.*
- *Sašo Tomažič is with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. E-mail: saso.tomazic@fe.uni-lj.si*

[1] Frame loss is a very important measure that strongly influences the quality of reconstructed speech. In Ethernet networks, frame losses arise primarily due to excessive delays, and so we will not discuss them in detail here (for details, see [6]).

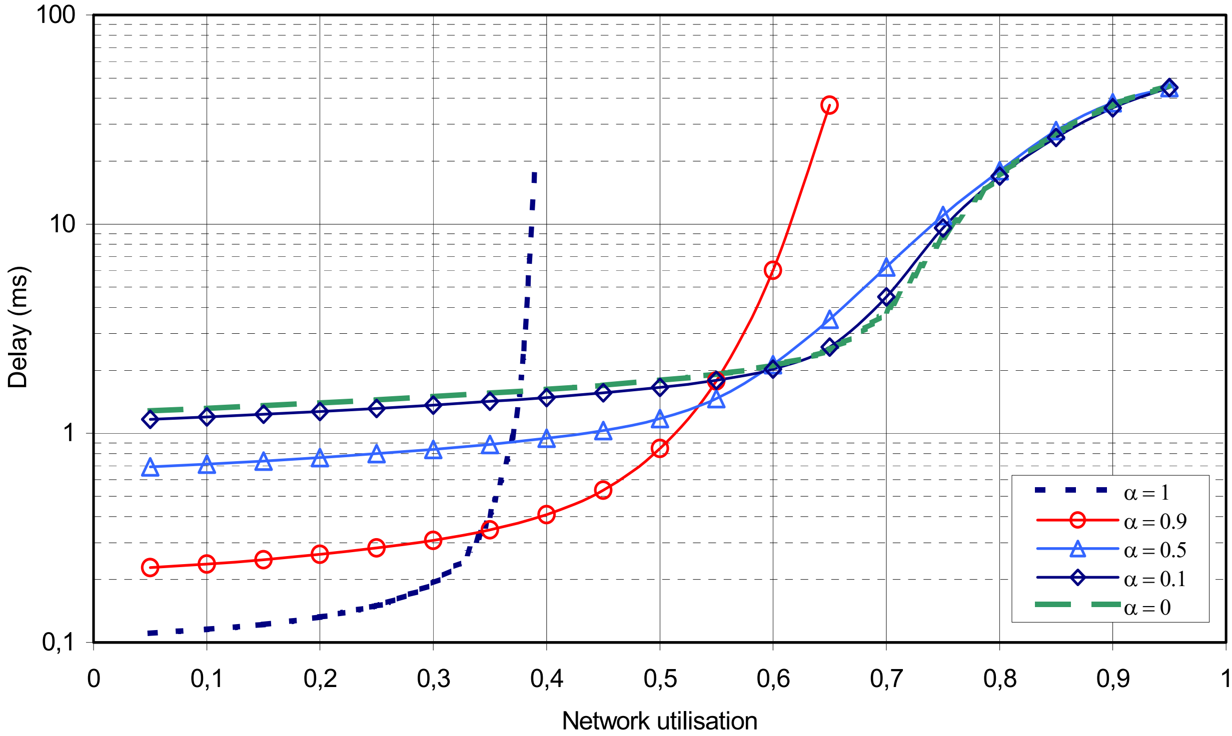

Figure 1: Average delay with regards to network utilisation and the proportion of short frames (α)

speech signal (7.5 ms) and frame coding/decoding (30 ms). The last value changes depending on the processing power of a given system, and could be lower (powerful system) or higher if the coder and decoder both need 30 ms to process a frame. In this case, the delay due to speech processing could grow to 97.5 ms.

4 Theoretical Analysis

A considerable amount of theoretical CSMA/CD network performance analysis is already available in existing literature. However, most is based on simplified models. The simplifications were mostly introduced due to the difficulty of mathematically modelling the operation of the network or simply to avoid dealing with to many parameters of the model. The following measures are particularly important in analysis of CSMA/CD networks:

- ***Average delay*** – average time from the first attempt to transmit a packet to receipt of the whole packet (collisions and retransmissions of collided frames can occur).
- ***Throughput*** – the proportion of the transmission channel bandwidth, that is used for the transmission of useful data, which includes the header and payload of the Ethernet frame.
- ***Channel capacity*** – maximum possible throughput for given parameters.

The most important measure for voice transmission is average delay, which we will use in the analysis of voice transmission over Ethernet networks.

The most common simplifications made in available analysis are:

- ***Packet arrival distribution*** – most theoretical analyses rely on a Poisson process, which does not give us the real state.
- ***Packet length distribution*** – most theoretical analyses assume simple distributions, such as exponential, bimodal, uniform or fixed packet length. The last in particular simplifies analysis, but usually consider only short packets, which give the worst possible result, or only long packets, which give the best possible result.
- ***Buffering and data flow control at higher protocol layers*** – usually neglected.
- ***Population of active transmission stations*** – mostly infinite population is assumed.
- ***Topology*** – mostly balanced topology with uniform (usually maximum possible) distance between stations on the transmission medium is assumed.
- ***Slotting*** – assumes that each transmission begins only at the start of an individual (imaginary) time slot. In real networks this is not the case.

Despite these deficiencies, theoretical analysis is most often the fastest route to answers about the operation and behaviour of a network.

4.1 Packet length distribution

The length of Ethernet frames depends on the applications running over the network. Given that most use the TCP transport protocol, we can expect that a considerable number of Ethernet frames on the network will be of minimum length (acknowledgements of TCP segments) – we denote them as short frames. Equally, we can expect that a considerable number of frames transmitting data will be of maximum length – we denote them as long frames. When a VoIP application is running, we can also expect a considerable number of short VoIP frames on the network.

We therefore carry out an analysis with the following simplifications:

- bimodal distribution of frame length (we have only two possible frame lengths, long and short),
- Poisson process of frame generation,
- infinite population of active transmission stations and,
- balanced topology.

An extensive analysis of delays on Ethernet network is given in [2]. For the purpose of our VoIP delay analysis, we will take advantage of the results derived for bimodal distribution of frame length.

Analytical results in [2] include a factor α, which represents the proportion of short frames (proportion of long frames is thus 1-α). We are given the expression for average delay of all frames with regard to the network utilisation and the factor α. The results for a 10 Mbit/s Ethernet network are shown in Figure 1.

When the network is loaded only with short frames (α=1), the turning point[2] of delay is very low at just over 30% network utilisation. With the lowering of value α, the turning point of delay shifts towards higher values of network utilisation and reaches 70% at α=0, when network is loaded only with long frames.

Notice that even a small number of long frames (α=0.9) can significantly correct the delay and shift the turning point of congestion towards a considerably higher network utilisation value of just over 50%.

If we assume that the transmission of voice frames represents only a small proportion of total traffic on an Ethernet network (α=0.1), and that other traffic mostly consists of long frames, then the delay to voice frames on the Ethernet network does not significantly contribute to total delay all the way up to 70% network utilisation.

When most of the traffic on the network belongs to TCP a long data frame is normally paired with a short confirmation frame, thus α is approximately 0.5 and turning point at 60% network utilisation.

From above, we can state that voice transmission over Ethernet networks is possible (especially when at least some long frames are present), but we can say nothing very specific regarding its characteristics, since too many simplifications were made in theoretical analysis. For more precise results, we have to turn to simulations or even better measurements on real systems.

5 MEASUREMENTS

In measurements on a real system, we were limited by equipment and the measurement procedure. We carried out measurements for three typical representative VoIP applications: PC telephone, independent IP telephone and gateway. For all three we measured VoIP delays on an unpreloaded network (network that is loaded only with VoIP traffic) and on a network preloaded with data traffic.

Measurements took place in two steps. In the first step, we measured the total delay between two VoIP terminals in an unpreloaded network. Here we actually measured the capacity of individual types of VoIP terminals. In the second step, we were primarily interested in the dependence of delay on the network utilisation. Network preloading was done with the help of a program that transmitted to the network data packets with a uniform length distribution between 64 and 1518 octets. Network utilisation ranged between 0% and 65%. The upper limit represented the maximum we could achieve with the available equipment.

[2] This point defines network utilisation value after which the delay starts to grow rapidly and quickly becomes unacceptable in terms of real-time voice transmission.

5.1 Unpreloaded network

Analysis of delays in an unpreloaded network was the basis for measurement in a preloaded network. Table 1 shows the structure of total delay D_{SK} between the transmission and reception terminal of the PC telephone. It only gives values for PC telephone for which we could measure the values stated (for the other two devices, this was not possible).

Delay component	Value
PC telephone (transmission) D_{PCO}	95
Network D_{IP}	~1
PC telephone (reception) D_{PCS}	396
Total D_{SK}	492

Table 1: *Structure of voice delay for the use of two PC telephones on an unloaded network*

With a PC telephone, we can measure the time D_{PCO}, which it requires to transmit a packet to the network. This delay comprises hardware delay (sound card and network card) and voice processing on the computer (Netmeeting software and TCP/IP stack of the Windows operating system). D_{PCS} is a delay on a receiver side and comprises of the same hardware delay and a jitter buffer delay. D_{IP} is a propagation delay on a physical transmission medium.

We can see that the network delay D_{IP} in an unloaded network is minimal, amounting in our case to approximately 1 ms. This delay increases in a loaded network, while the delays D_{PCS} and D_{PCO} in principle remain the same.

5.2 Preloaded network

Due to the limitations of measuring equipment, we were not able to directly measure delay in the transmission of packets D_{PCO} and D_{PCS} in a preloaded network. It is necessary therefore to emphasise that the values given in Table 1 for D_{PCO} and D_{PCS} change somewhat with increasing load, since they are also dependent on the processing load of the computer on which the PC telephone operates. This introduces a certain degree of uncertainty into the calculation of delay on a loaded network D_{IP}, which we obtain from the equation

$$D_{IP} = D_{SK} - D_{PCO} - D_{PCS} \quad (1)$$

Using our measurement system, we could on a preloaded network only measure delays D_{SK}, while the values of delays D_{PCO} and D_{PCS} are taken from measurements on an unpreloaded network.

We carried out several series of measurements on a preloaded network. The network preloading procedure did not allow uniform load intervals; instead these intervals changed somewhat. This can partly be ascribed to variable process loads on the computer, partly to the principle of

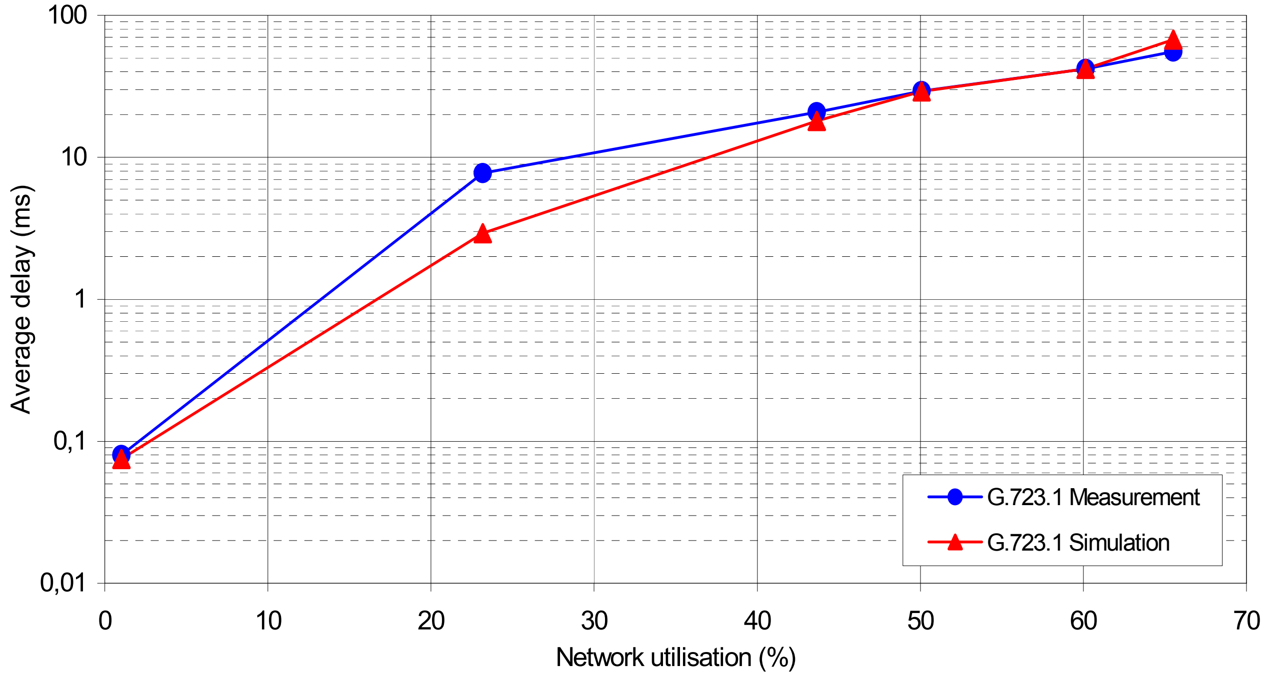

Figure 2: Comparison of delays on a loaded network between PC telephones and simulation results

operation of Ethernet networks (influence in higher loads), and partly to measurement error of the network analyser.

6 Simulations of measurement system

In addition to measurements, we simulated the previously described measurement system. Here we wanted to approximate as closely as possible the scenario of measurement on a real system. To this end, we simulated the topology of the network of the measurement system and adjusted traffic-generation in simulation to the traffic generation in measurements. Likewise, we simulated the protocol stack for transmission of VoIP data and data of load traffic in the manner carried out in the measurement system. The purpose of this exercise was to test how closely to real results can we get with simulations.

6.1 Comparison of measurement and simulation results

Since the parameters of the simulated system were chosen as closely as possible to the measuring, it was to be expected that the simulation results would be close to the measurement results. The measurement system only allowed us to measure the total delay D_{SK}, while the simulations provided the network delay or D_{IP}. To avoid the uncertainty in the calculation of D_{IP} from equation (1), we decided to use the minimum logarithmic deviation procedure. Under this procedure, we seek the minimum of function $g(\Delta)$, where n represents the number of samples of functions f_1 and f_2, and Δ is the difference between the two functions. In our case the function f_1 represent measured values of D_{SK} and function f_2 simulated values of D_{IP}.

$$g(\Delta) = \sum_n \left| \ln(f_1(n) - \ln(f_2(n) - \Delta) \right| \qquad (2)$$

Figure 2 shows the network delay against the network utilisation. It contains curves with the results of measurement and simulation for PC telephone and voice coder G.723.1. Looking at Figure 2 in more detail, we see that the results of measurement and simulation match fairly well. This particularly applies to higher network loads. The differences between results are in the range of a few milliseconds, which can be considered to be a good approximation.

This comparison shows that simulations can sufficiently predict the operation of VoIP applications. This opens up the possibility of forecasting their behaviour under various conditions even before they are implemented, as well as detailed study of various operating scenarios before system modifications.

7 Discussion

Due to limited space, measurement results given in this paper are in a compact form. Detailed results can be found in [4], [7] and [8]. Theoretical results are based on certain assumptions and simplifications that are usually introduced to make the model mathematically manageable. The great value of theoretical calculations is in fast results, providing general relations between network parameters and setting benchmarks for other methods (simulations and measurements). Theoretical results shown in Figure 1 emphasis the properties of the voice coder G.723.1, so that they can be compared with results of simulation and measurement.

The second phase included measurement on a real system and its simulation. Comparison between them is shown in Figure 2. In the simulation scenario, we did not introduce simplifications, and so direct comparison with the results of theoretical analysis is not possible.

The key question, which also led to this work, is: “Is the Ethernet network able to transmit interactive speech under recommendation G.723.1?" And if so, to what extent and under what conditions. For a complete answer, we have to set an upper limit for acceptable delays for speech; these are given in section 2. Here, we must realise that this is the delay on the entire transmission path, which is discussed in section 3 and in more detail in [5]. If we deduct from the upper limits of delay (150 and 400 ms) the voice processing delay, which is somewhere between 67.5 and 97.5 ms, and we add some 10 ms delay due to the jitter buffer, we are not left with much for network delay. Measurements on an unpreloaded network in section 5.1 and the values in Table 1 indicate that total measured delay DSK is at or even above the upper limit, where the network delay is only 1 ms.

If we are a little less demanding, and an extra 10, 20 or 30 ms is not very important, then we can say that voice transmission will be satisfactory up to around 50% network utilisation. But if we want to keep the network delay below 10 ms, then the network utilisation should be below 40% (for both see Figure 2)

With regard to the comparison of measurement and simulation results in Figure 2, we can say that, when taking account of all conditions and situations on the network, the simulations come close enough to real results. This means that we can simulate different topologies, settings and scenarios of VoIP networks even before they are physically installed and the equipment purchased.

8 Conclusion

Comprehensive consideration of conditions on an Ethernet network and investigation of the possibility of real-time voice transmission provided answers to most of the questions posed. We learned that theoretical analysis – due to the simplifications introduced – is mainly suitable for studying general network conditions and, in our particular case, determining upper bounds of delay. We can obtain much better results with simulations that use fewer simplifications and better approximate a real network, however the only way to obtain exact result is through measurements on a working system.

This does not imply that that theoretical analysis and simulations are unnecessary. Each serves its purpose, and they are an irreplaceable aid in planning of systems. It is only important that the correct assumptions are used; otherwise the results could be too far from the reality. Which assumptions are correct we can only learn through measurements on a working system. All three methods thus complement each other. Their appropriate use can save us much discomfort and many unpleasant surprises.

References

[1] David R. Boggs, Jeffrey C. Mogul, Christopher A. Kent; *Measured Capacity of an Ethernet: Myths and Reality*, ACM SIGCOMM Symposium 1988

[2] Joe Parry; *Two Models of Ethernet Networks*, http://www.esat.kuleuven.ac.be/~h239/model/ethernetmodel.pdf

[3] Jia Wang, Srinavasan Keshav; *Efficient and Accurate Ethernet Simulation*, Proceedings of the 24th Conference on Local Computer Networks, October 1999

[4] A. Vugrinec, *Prenos govora preko omrežja z internet protokolom*, Master's dissertation, Faculty of Electrical Engineering, University of Ljubljana, June 2001

[5] Anton Kos; *Prenos govora preko podatkovnih omrežij*, Master's dissertation, Faculty of Electrical Engineering, University of Ljubljana, April 1998

[6] Anton Kos, Borut Klepec, Sašo Tomažič; *Techniques for Performance Improvement of VoIP Applications*, Melecon 2002, pp. 250-254, May 7-9, 2002, Cairo, Egypt

[7] Anton Kos, Sašo Tomažič; *Teoretična analiza in simulacije prenosa govora preko krajevnih omrežij Ethernet*, Research work of postgraduate students in Slovenia – One Science, e-proceedings, Ljubljana: Association of Young Researchers of Slovenia, 2002, pp. 141-151

[8] Anton Kos, Sašo Tomažič; *VoIP over Ethernet – Theoretical Analysis and Simulation*, SoftCom 2003, Split, October 2003

[9] ITU G.114: *One-way transmission time*, International Telecommunication Union, Geneva, 05/2000

[10] Martin Rowe, *Measure VoIP Networks for Jitter and Loss*, Test & Measurement World, December 1999, str. 29 – 38

[11] Viktor Matic, Alen Bazant, Mladen Kos, *Optimal delay buffer for VoIP applications*, Proc. of 14th European Simulation Multiconference (ESM2000), Ghent, May 23-26, 2000, pp.521-525

[12] Thuan Nguyen, Ferit Yegenglou, Agatino Sciuto, Ravi Subbarayan, *Voice over IP service and performance in satellite networks*, IEEE Communications magazine, March 2001, str. 164 – 171

[13] Harri Marjamäki, *Performance evaluation of an IP voice terminal*, Proceedings of the Fifth IFIP Conference on Intelligence in Networks, Asian Institute of Technology, Thailand, November 1999

[14] Jiankun Hu, Noel Low, *Investigation of Telephonic Voice Traffic Network Performance on Integrated Wireless LAN and Wired Internet*, IEEE International Conference on Telecommunications, Romania, Bucharest, 4-7 June, 2001 (IEEE ICT 2001)