ISCA Archive
http://www.isca-speech.org/archive

European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

A MODEL FOR ASSESSING SUBJECTIVELY THE EFFECTS OF DELAYING SPEECH PACKETS IN A PACKET SWITCHED NETWORK

B.M. LINES*

ABSTRACT

An initial hardware/software model has been developed to assess the ability of a local area network (LAN) using the Technical Office Protocol (TOP) to carry interactive speech, as part of a MAP[o]/TOP network in a Computer Integrated manufacturing (CIM) environment.

INTRODUCTION

The modern generation of manufacturing industries are currently adopting the CIM strategy as a means of optimising their operation [Ref1]. Two important aspects of this strategy are the automaticon of the office and the manufacturing processes. As a result, various computer networks have been suggested to support these two activities. Current standards emerging [Ref2] indicate that LANs will provide the necessary facilities for satisfactory interconnection. In addition to the transfer of data between workstations and processers a need has been established for verbal communication between users of the networks.

LAN TOPOLOGIES

Two distinctly different LANs are emerging as standards in their own right, they are:-

Office Automation:	Bus structure employing CSMA/CD‡ with Technical Office Protocol (TOP) - IEEE 802.3
Manufacturing Process:	Token passing bus using the Manufacturing Applications Protocol (MAP) - IEEE 802.4

DELAY ISSUE ON PACKET SWITCHED LANs

The major issue in the transmission of speech over LANs is that the delay of any one speech packet must be minimised [Ref3,4,5 and 6].

The amount of delay a packet suffers is a function of:-

(a) The network topology
(b) The level of traffic offered to the LAN.

At this point in the research programme investigations were carried out into the subjective effect on human conversation of delaying speech packets and returning them out of order, and thus obtain a measure of the delays involved. To achieve this aim a model was designed and built.

‡ CSMA/CD - Carries Sense Multiple Access/Collision Detection.
* Dept. of Electrical & Electronic Engineering, Plymouth Polytechnic.
o Manufacturing Applications Protocol.

10.21437/ECST.1987-174

STATISTICAL NATURE OF DELAY ON CSMA/CD N/Ws

To ascertain the levels of acceptability the model was designed for two users such that the delays in either direction could be controlled separately. In addition, the nature of the delays had to mirror those likely to be encountered on typical MAP/TOP networks. The initial investigations were concentrated on the CSMA/CD TOP networks where the nature of the delays are likely to cause more distortion of the speech than on the Token Passing Bus employed on the MAP LAN where delays are more easily controlled.

Statistical information obtained from other CSMA/CD [ref 5] on the length of all types of delays provides a typical distribution.

The shape of the distribution is explained as follows:-

(i) There will be a small fixed delay as a result of the time taken to seize the network assuming no other traffic offered. This will include the propagation time of the seizing packet to and from the termination at the end of the cable [ref 4].

(ii) As the traffic increases variable delays will occur in addition to the fixed delays defined in (i) above. The nature of these delays will be a function of the number of stations in contention, and the size of the data packets being employed by non speech users. [ref 4] [ref 5] [ref 6]. In addition the "back-off" algorithm employed will also influence the length of the variable delays.

DELAY MODEL

An explanation of the hardware-software model of the packet switched network built and used innthe subjective testing follows:-

The digitising of the speech follows the usual pattern of low-pass filtering with a cut off of 3.5 kHz and sampling at 8 kHz with 8 bits per sample. The only departure from the norm was the use of a switched capacitor type filter to achieve a faster rate of cut-off and hence minimise aliasing. The analogue to digital convertors were pods linked to a versatile interface adaptor (VIA) connected to the BBC B via the 1 MHz bus to minimise processing time.

The 8 bit bytes arriving from the A/D pod were written into the memory of the BBC via one part of the dual VIA and the 1 MHz bus. The reading-out process to the D/A port of the pod was under operator control. This control was:

(a) fixed delay value from 0 - 1 sec in steps of 10 msec from a switch control box - separate for each direction of transmission

and/or (b) varible delay under operator control such that the standard deviation of the distribution (a truncated Gaussian) could be altered from 1 msec to 400 msec - separate for each direction of transmission.

The variable delay was produced by using a white-noise generator with amplitude shaping filters followed by A/D conversion. The numerical values of the samples obtained were used, after suitable manipulation, to control the delay value imposed prior to reading out the speech packet value.

A third BBC B was used to show the delay values and produce a histogram on the display as well as a visual numerical value of the delays with a 0.5 sec refresh time, - thus giving an indication of the standard deviation as well as the statistical shape of the delays.

RESULTS

Fixed Delay

Investigations were carried out on the tolerance of subjects [ref 7] to a fixed delay using non]contextual subject matter. It was felt that if levels of acceptability could be obtained in this way then such a system would be easily intelligible for contextual conversation, which is its normal mode of operation.

Results showed that the important parameter is the total delay i.e. the time taken for the initiator of a sentence to receive some acknowledgement of a statement. This was termed LOOP DELAY . A figure of 250 msec was found acceptable by most users (over 70%), whereas less than 70% of those tested found 500 msecs unacceptable. An additional factor worthy of note was the apportionment of the delay i.e. Total loop delay = delay channel A + delay on channel B.

Subjects tested found in the case of large delays that there was the well documented problem of talkers repeating their initial statements having not received initial responses. As a result considerable time is lost in users acquiring the "half duplex"‡ skill when talking over heavily delayed network.

Variable Delays

It must be pointed out that as investigations were being made into the effects of raw delay on a TOP LAN, no attempt was made to reorder the speech packets. Thus if the (n+1)th packet is delayed less than the nth packet, then when presented for reconstitution, the (n+1)th packet will appear for reconstitution <u>before</u> the nth. As a result distortion of the speech occurs.

Results indicated that the application of variable delay on a fixed delay had very serious effects on speech quality, using non-contextual speech tests. The move to contextual test material to improve intellgibility had little effect.

From tests conducted so far a standard deviation of 2 msec on any fixed delay not exceeding 250 msec is intelligible but only just. Any marginal increase in standard deviation above 2 msec renders the speech completely unintelligible. There is a sharp knee in the curve which is much sharper than for the fixed delay. An explanation for this can be found by investigation of the delays being experienced and the

‡ half duplex skill is to speak-then wait for a reply before speaking again.

effect on the arrival of the samples and the reconstitution of the analogue waveform. A delay "hit" of 2 msecs will contain 2 msec/125 µsec samples = 16 samples. This implies that 16 speech samples will have been misplaced when reconstitution takes place. It would therefore be expected that the speech quality is significantly impaired. For standard deviations less than 2 msec the level of acceptability improves dramatically.

CONCLUSIONS

From the results so far achieved there are a number of recommendtions that can be made to improve the quality of speech on TOP LANs, they are:-

1. Minimise the fixed loopm delay to less than 250 msec.

2. Ideally keep any variable delay as low as possible - less than the 2 msecs standard deviation.

<u>or</u>

3. Reorder the speech packets at the receiving end by buffering the incoming speech packets (i.e. insert a fixed delay that <u>masks</u> the variable delay).

4. If speech packets are delayed by more than 250 msec, then ignore those missing packets and interpolate between successive samples (by low pass filtering or software techniques).

Under very heavy loading, the darta traffic carried on a CSMA/CD network will cause the system to crash when it reaches 30% of its theoretical maximum bit rate [ref 5]. This knee is very sharp and should not be approached. Another approach to improving the speech carrying capacity of a CSMA/CD LAN is to employ a priority protocol which will allow speech packets fast access to the LAN. Research is currently underway in this area. In addition, investigations are being carried out to improve the efficiency of speech transmission by placing multiple speech packet in a frame. The subjective impact of loosing multiple successive packets is under investigation, as are the ways of minimising the effect of such a loss.

REFERENCES

1. White Paper - Industrial and Commercial Automation - 1985.
2. a) IEEE 802.4/ISO 8802.4 - Token Passing Bus Medium Access Control.
   b) IEEE 802.3/ISO 8802.3 - CSMA/CD (ETHERNET).
3. J.E. Flood & C.C.F. Tucker, 'The Advantages of Packet Switched Speech', University of Aston, Birmingham, 1985.
4. A.S. Tanenbaum - Computer Networks, Prentice Hall 1981.
5. Dunlop Rashid - Speech Transmission Capacity of Standard Ethernet Systems, IERE Journal, Vol.55, No.3, pp 119-122, March 1985.
6. John G. Gruber, Delay Related Issues in Integrated Voice and Data, IEEE Trans. on Comms. Vol. Com-29, No.6 June 1981.
7. Richards, D.L. 'Telecommunications by Speech' - Butterworth 1973.