ANALYSIS OF A USER COMMUNICATION PATH IN A STORE AND FORWARD NETWORK.

GONZALEZ SOTO, O. , GUTIERREZ, D. , MARTINEZ MIGUEZ, L.

Centro de Investigación de Standard Eléctrica,S.A. Madrid. Spain.

ABSTRACT
A user communication path is defined as the subnet­work formed by the user equipment at source and destination, plus the chain of nodes and transmission lines which provide their interconnection in a store-and-forward network.

A user communication path is modeled by a network of queues that represent user behaviour, node processing and the transmission links for each direction of transmission. The paper describes the application of this modeling technique to single virtual calls in a packet switching network, and to multiservice calls accessing an integrated services digital network via the D-channel. Results such as the throughput and delivery delay of the path are given as a function of protocol and traffic parameters. Average values are obtained analyti­cally; detailed sensitivities and distributions are obtained by simulation.

1. INTRODUCTION
The paper defines a user communication path in the sense of the complete chain of user and network equipment, either for single or multiple users. The purpose of our present study is to model the traffic in the complete user-network path. Because of the presence of user network feedback in a store and forward network with flow control, emphasis is placed on user behaviour, processing at the nodes, and the protocol parameters involved.

Initial works on communication paths (1,2,3) considered unidirectional channels within a switching network and infinite capacity nodes. In (4,5) congestion control mechanisms are included in the path and the works of (6) analyze the collection of paths which compose a network. A single chain with explicit consideration of the processing times at the nodes is analyzed in (7) to derive the associated efficiencies. Other works (8-10) also consider several types of paths under the more or less common assumption of packet independence.

The path definition introduced in the paper was initiated in the modeling of a packet switching network with X.25 boundaries, specifically the DN-1 network (11), by four hierarchical levels (12) under the name of NEPER-HIM. They are summarized in the Figure 1. The user communication path model provides the end-to-end grade of service and performance, assuming semi-independence of the details associated to the lower levels (module and node) but using their results.

Due to the extension of X.25 or X.25-like interfaces to many networks, the model application becomes valid in the much broader area of Telematic, mixed and value added networks. The complexity of mixed networks of packet and circuit type, for example ISDN (integrated services digital networks) makes this type of hierarchical modeling even more attractive. In particular the user communication path in an ISDN was applied for the grade of service derivation, either for complete paths or for sub-paths.

Two examples of this application are described in the paper:

- Single virtual calls associated to the path in a packet switching network from a terminal to a host computer
- Multiple virtual calls from a multiservice installation that accesses a packet switching gateway via a D channel.

The user communication path is modeled by a network of queues that are represented with a high degree of detail in the simulation approach and maintaining the packet dependence, while less detail is used in the analytical approach. The re-
A communication path node is identified as such when it has a store and forward capability. If other network nodes are crossed in a transparent way to the flow, i.e. a circuit switching node of a telephone network, they are represented by the slowest transmission link to the next store and forward node in the physical path. Origin and destination equipments are considered as the first and last path nodes respectively.

A basic characteristic of the path in a store and forward network is the flow control existing in the communication, which includes a strong feedback between the user and the network. The information is accepted as a function of the acknowledgments of previous packets, window size, etc.; the arrival law of packets and the holding time of calls are a consequence of the interaction between the user, the network and the protocols.

The context of the user communication path model as the third level of the NEPER-HIM package may be expressed in terms of the following general assumptions:

a. The user side of the communication path is represented with a degree of detail that includes user behaviour, with parameters such as the user thinking time, net information generated, processing time at user equipment, reaction time, etc.

b. The network side of the communication path is represented for those modules and resources which are directly used by the communication under study, while the effect of other modules and traffic is considered in a simplified way by a statistical interference traffic.

This is one of the basic distinctions with the fourth level of the NEPER-HIM package: the global network model has a reverse representation with more detailed consideration of the transport part of the network and simplified statistical modeling of the network periphery. This type of representation for the user communication path is quite logical in an actual network, since thousands of paths interact in the central part of the network, each path having a very slight effect. This allows the semi-independence assumption with regard to the global network model. The correspondence of this modeling with the functional operation is particularly close when path-oriented routing (fixed, random, etc.) is used in the network.

The objectives to be sought, and results attained, for the user communication path can be summarized as follows:
- Analysis of the user-network interdependence
- Determination of throughput of a path for each configuration and design
- Analysis of the traffic-protocol interactions in the path due to the flow control mechanisms
- Investigation of arrival laws of the packets within the same path, as well as the holding time distribution of calls
- Capacity of the path in terms of simultaneous communications, number of terminals, etc. for a given state of the network
3. ANALYTICAL MODEL.

The problem that we face in this section is to solve analytically the throughput and network crossing time of a communication when two ends are connected. When the problem is treated in an analytical way, the main objective is to give a quick result of the end-to-end performance between users, but with more restrictive assumptions (for example in the source generator and flow control algorithms) than the assumptions used in the simulation approach. Comparison with the results derived from the simulation model give a measure of the simplifications made in order to attack the problem analytically, and thereby render it more tractable.

For the analytical study of a user communication path we only treat the end-to-end flow control. The existence of a window flow control allows to transform the initial open network of the user communication path into a closed one.

In the Figure 6 we represent the resources that are implicit in this user communication path model as a tandem network. Thus we represent the processing nodes, transmission links, and source/sink processing in a simplified way.

Based on the general assumptions of the previous paragraph, we have made the following additional considerations:

1. It can be taken into account that the interfering traffic due to other calls reduces the maximum transmission link capacity \( \mu_1 \) by the average arrival rate of the interfering traffic. This means that if the maximum transmission capacity of a link \( i \) is \( \mu_i \) and the interfering traffic arrival rate at link \( i \) is \( \lambda_i \), the net maximum transmission capacity (5) will be: \( \mu = \mu_i - \lambda_i \).

2. The transmission links are represented by a first-in-first-out queue that only treats packets/acknowledgments with exponentially distributed transmission time. Exponential distribution is a good approximation as high as the interference is.

3. The source is represented by a packet interarrival time, an average packet length and a processing time. The sink has a processing time; it generates an acknowledgment without priority and send it in the reverse direction.

4. The crossing nodes are represented by a pure time delay with phase-type distribution.

All the assumptions made up to now are extended to the multiple virtual calls traveling by the same user communication path.

Based on the previous assumptions, we obtain the above mentioned closed queuing network with a finite population analytically tractable. For the calculation of mean queue sizes, mean waiting times and throughputs in closed queuing networks with product-form solution (16) without computing the normalization constant, the mean value analysis has been introduced (17). This technique computes the performance measures of networks with product-form solution based on the two following principles:

1. The relation between the mean waiting time and the mean queue size of a service center with one less customer (18, 19).

2. Little's law applied to each service center and throughout the chain (20).

A good characteristic of the mean value analysis in comparison to the corresponding normalization constant method is that in addition it provides an easy understanding from an intuitive point of view (21).

In the following we derive the algorithm for the calculation of average performance parameters for a multiple call communication path. We introduce, then, the following notation:

- \( C \): Number of existing calls in the communication path
- \( W_C \): Variable for the window size of call \( c \), \( c=1,2,\ldots,C \)
- \( T_1 \): Mean service time of a data packet of call \( c \) in transmission link \( i \)
- \( T_\delta \): Mean delay time of a data packet of call \( c \) in node \( j \) for pure delay nodes
- \( T_\delta' \): Mean service time of a data packet belonging to a call \( c \) in a FIFO node \( j \)
- \( I \): Number of transmission links in the communication path
- \( Q \): Set of FIFO queues in the C.P.
The reason for differentiating the existing calls above expressions will be simplified and the number size per call, different classes of traffic, different call phases, etc. and forward application, such as different window values will be obtained.

Following the notation introduced above, the first principle of MVA can be expressed as:

\[ n_k^1(\cdot) = n_k^2(W - e) \]

and relating it with the second principle (Little's law) we have:

\[ t^c = \frac{n^c}{\lambda^c} \]

for pure delay nodes.

\[ t^c = T^c_k(1 + \sum_{i=1}^{\lambda^c} n_k^i(W - e)) \]

for FIFO or Processor Sharing nodes.

\[ t^c = T^c_k(1 + \sum_{i=1}^{\lambda^c} n_k^i(W - e)) \]

for transmission links.

\[ \lambda^c = W^c / \sum_{k=1}^{I+J} n_k^c \]

\[ n_k^c = \lambda^c t_k^c \]

Starting at the empty state \( n_k^c(0) = 0 \) and applying that in an iterative way, the performance values will be obtained.

The high degree of interaction produced at the user side and the network nodes where the flow control is executed originates complex packet arrival laws. In order to analyze their effects as well as to produce more accurate and extended results like distributions, simulation has been used.

The simulation model is mainly based on the assumptions of paragraph 2 without special constraints. The specific characteristics of the simulation are summarized:

a) A two stage generation is done with Markovian generation of communications or calls by the call generator (CG) at the user side and a second packet generation (PG) which originates a packet flow within the entire call with complete correlation and dependence on the flow control handling and statistical network behavior. The user reaction time to each acknowledgment and thinking time are represented due to their impact in the path and to derive user/network interaction.

b) The specific distributions of data packet per call, packet sizes for each traffic class are taken into account at the user generator. This allows the detailed representation of the actual statistical values for all the traffic mixes. Within the data transfer phase, forward and backward information traffic may be represented as well as different sessions.

c) The packet flow at the path originated by the different procedures of the flow control, may be represented with the adequate degree of detail (14). The various implementations of the flow control algorithm with their facilities such as local credit handling, anticipation, etc. are easily implemented at the nodes with that facilities.

d) The functional representation of the nodes is made with an intermediate level of detail for any interconnection topology. Each node is mainly implemented by the set of the following five queues:

- The packet processing/switching queue (PQ) which compounds the behaviour of the processing associated to level 3 of the protocol for each packet class.
- The forward and backward input queues (FIQ, BIQ) which represent the processing associated to the frame reception.
- The forward and backward output queues (FOQ, BOQ) which represent the processing of the frame sending plus the link transmission itself.

The figure 7 represents a diagram for the queues of typical nodes of the path. The path nodes that have specific functions like the flow control handling at the network gates or like the user generation at both ends have the same basic structure summarized above with additional internal feedbacks.

e) The processing times at the referenced queues and the delay type services may be associated to each packet type, call class and node for any distribution. This input values are provided by the lower models of the NEPER-HIM hierarchy.
A markovian generator of interference traffic (IG) is provided per queue of the network side of the path in order to represent the rest of the network traffic in a simplified way.

The whole set of objective results given in the paragraph 2 are covered with this simulation model. Specially those related with the GOS are defined:

- Packet delivery delay or time between the input of the first bit of the packet at the first processing queue of the network and the output of the last bit of the same packet at the last processing queue of the network.

- Round trip delay or time between the output of the last bit of forward packet from the user equipment and the input of the first bit in the associated backward acknowledgment at the user side.

- Total call holding time, between the generation of a call request packet and the reception of the clear confirmation packet.

5. APPLICATION CASES.

Two application cases, one for single virtual calls and other for multiple virtual calls, are summarized as examples of the results that the defined communication path model may provide.

5.1 Single virtual call in an X.25 packet switching network.

The user to host path associated to virtual calls in the path of fig.3 has been analyzed by the analytical and simulation tools for the following set of parameters:

- Number of path nodes from 3 to 7.
- User line speeds from 2.4 kb/s to 48 kb/s.
- X.25 protocol at both sides with a standard maximum packet field length of 1024 bits.

- Interactive traffic having an average of 2.5 packets per call and uniform packet size with 75 octects in average.
- Batch traffic having 114 full packets/call.
- Window sizes from 1 to 8 with normal end-to-end significance of the flow control and other case with use of local credit.
- Delay and processing times at the different modules involved in the path obtained from the module analysis (MODA) of NEPER-HIM applied to the modules of the DN-1 network.

The fig.8 gives the end to end delay and some of the components as a function of the window size obtained by the analytical and simulation models. The configuration case corresponds to five path nodes at nominal loads and with a user line speed of 9.6 kb/s. The difference for the end to end delay between both approaches is not high (less than 15%) due to the fact that the bigger errors in some components (up to 70%) in the processing at

**FIG.7 - SIMULATION MODEL FOR THE USER COMMUNICATION PATH**

![Simulation Model Diagram]

**FIG.8 COMPARISON OF DELAYS**

![Comparison of Delays Chart]
The first node) are not always in the same sense. The delay components of the nodes where the flow control is implemented exhibit higher discrepancy between the analytical and simulation approach. The reason is the strong influence of arrival laws and processing times in that nodes where the independence and Poisson arrival assumptions become less valid than in the intermediate nodes of the network. Then care has to be taken with the validity of the previous assumptions according to the objectives of the study.

The figure 9 gives a comparison of the call holding times for interactive and batch classes of traffic as a function of the user line speed and for several number of path nodes and window sizes. The effect of user line speed is important for low speeds while for the highest ones the number of stages and window size are dominant.

The figure 10 gives the variation of typical components of path delay as a function of the user line speed. It is derived the interaction between the user and network sides of the path. For small components of user line delays there is an increase of the network side components due to the quick arrival of packets to the network resources and vice versa.

The figure 11 gives the total communication call holding time and delivery delay for a data packet as a function of the path load (assuming homogeneous load at the different path stages). Comparison is made between the use or not of local credit at the protocol. The result emphasizes the effect of the credit use in a batch class of traffic while slight difference is appreciated for an interactive class.

The distribution of delivery delay for data packet of interactive or batch traffic classes is given in figure 12. This is one of the typical grade of service associated values to be used for the path dimensioning and selection of protocol parameters.

5.2 Multiple Calls in a Multiservice Subscriber

We analyze in this paragraph the communication path shared by the simultaneous virtual calls originated by the terminals of a multiservice subscriber in an integrated environment. The type of integration is of an intermediate level in such a way that the services are provided through a gateway which ac-
cess a packet switching network. The figure 5 represents the analyzed stages.

The multiservice subscriber owns a D channel of 16 kb/s which carries the signalling information associated to B channels of 64 kb/s and also may carry additional information belonging to services like telemetry, videotext, teletex, etc. assuming that priority is used for the signalling.

In order to analyze the communication path of multiple multiservice calls, we have selected an homogeneous case, i.e. all the simultaneous calls are originated by terminals of the same service type with the same traffic and flow control parameters (average volume of information per call, window size, etc.). Within the future services we choose for the analysis the teletex service due to the fact that it has an important quantity of information to be interchanged and allows the channel analysis in the high range of load.

Typical values used for a teletex calls were 50 kb per page and an average of 2 pages per call with uniform distribution.

Due to the low volume of information produced by the signalling traffic we consider its impact on the teletex service to be negligible.

The figure 13 represents the resultant gross throughput of the communication path as a function of the number of simultaneous teletex virtual calls for different values of the window size. It may be observed how the highest window sizes saturate the channel with less number of simultaneous calls (between 3 and 10). Observe that the number of terminals associated to a D channel may be higher according to the traffic intensity per terminal and coincidence of busy periods.

Figure 14 gives the average access time of a packet as a function of the number of simultaneous virtual calls for different values of the window size. The packet access time considers the elapsed time between the entry to the network termination and the end of its processing by the gateway.

Defining the power (15) of a communication path as a relation of the path throughput to the average packet access time, the figure 15 represents the power as a function of the number of simultaneous teletex calls for typical values of the window size. It is clear that the high windows are optimum for low number of simultaneous calls, while the low windows are for the high traffic demand. The window size of 2 optimizes the path use for a wide range of simultaneous calls.

The figure 16 represents the packet access time distribution of the basic service, obtained by the simulation model, for three different numbers of the simultaneous teletex calls.

This probability distribution joint to the channel throughput and the power of the path provide consistent criteria for the dimensioning of the various configurations of the D channel.
REFERENCES


6. CONCLUSIONS.

The context of a communication path in a general store and forward network has been defined with the inclusion of user equipment behavior in the modeling and using the performance results provided by previous analysis of lower hierarchical levels.

The simulation tools developed for the path analysis where found particularly appropriate for the investigation of traffic-protocol and user-network interactions as well as for the end-to-end grade of service derivation. The applied analytical tools based on the mean value analysis seems to be a useful procedure for the evaluation of average performance values.

Two applications were briefly exposed for single virtual calls in a packet switching network and multiterminal calls in an ISDN environment. From the comparison of different simplificative assumptions it is derived the need for consideration of different packet classes, arrivals in groups and packet dependence in that sectors of the path where flow control is exercised or multiple generation appears.

Finally, the scope of this path modeling seems to be quite large for actual data networks and sub-networks like packet switching, CCITT no.7 signaling, ISDN and in general mixed networks which follow the Operating System Interconnection layered structure.

REFERENCES

Q.1 (W. Dieterle)

In contrast to the results published in my own paper (session #3.3 paper #6) your curves in Fig. 8 show an increase in delay when enlarging the value of the window size. Can you give a short explanation of the behavior of the delay-time curves due to the window size?

A.1 (O. Soto, D. Gutierrez, L. Miguez Martinez)

After a quick look to both papers, it is doubtful if both results may be compared due to the quite different modeling assumptions involved:

- Mainly you modeled the level 3 as a two adjacent stages with infinite source arrivals and with packet independence.
- We modeled the level 3 as the full path through the network and including the end user behavior implying:
  - Finite source effect
  - Time correlation associated to flow control
  - Bunch arrival effect associated to the packets when the window is open

These three assumptions and mainly the last one produce an increase of packet delay for higher windows as is described in the paper. In parallel, higher windows reduce "call holding times" due to the less number of packet-ack cycles.

Finally, I emphasise the importance of assumption adequacy to each problem mainly in this type of new application. We may discuss more details if you are interested. Thank-you.