A SIMULATION-BASED PERFORMANCE STUDY

OF NETWORK-LAYER RELAYS

Authors:

MÁRIO BERNARDES, EDMUNDO MONTEIRO, and FERNANDO BOAVIDA

<u>Affiliation and Address:</u>

Laboratório de Informática e Sistemas

Universidade de Coimbra

Urb. Quinta da Boavista, Lote 1, 1º

P-3000 COIMBRA

PORTUGAL

Tel.: +351 39 724275

Fax: +351 39 724266

SUMMARY

Functional standardization activities in ISO and in regional workshop are currently addressing network layer relays for the interconnection of several types of subnetworks. In this paper the authors present a simulation-based study of two different approaches - the service relay approach and the protocol relay approach - for LAN/WAN interconnection. The study draws some attention on several aspects of network layer relays, and can be used as a basis to predict and evaluate their general performance characteristics. In order to study and compare the two different relay types, a number of simulation studies were carried out, using several traffic conditions and relay models. The obtained results show, on one hand, that both relay types are useful, and that more attention should be given to the service relay approach by standards developers and relay implementors. On the other hand, the results highlighted the weight of the relays internal architecture in their overall performance, and pointed out the need to carefully study relay configurations for high speed networks interconnection, that can be used without being the limiting element of internetworking solutions.

KEY WORDS: Service relay, protocol relay, network-layer relays, internetworking

1. INTRODUCTION

The generalized use of local area networks (LANs), and the need for information interchange between users connected to different types of LANs, possibly located in geographically distant places, has lead in the last few years to the development of LAN interconnection solutions based on wide area packet switching data networks (PSDN WANs) (Schepers *et al.*, 1992a).

Having in mind that LANs may use different medium access methods, have little need for network layer functionality and, consequently, use addressing schemes that are different from those used in WANs (EWOS, 1990a; Schepers *et al.*, 1992a), we can easily understand that there are several important issues in LAN/WAN interconnection, that can be dealt with in a lot of different ways. To prevent incompatibilities between different interconnection solutions, there is an ongoing functional standardization activity in ISO aiming at the development of International Standardized Profiles (ISP), that benefits from the harmonized input of regional workshops (EWOS[1], NIST OIW[2], and AOW[3]). This will, hopefully, lead to standardized and compatible implementations of relays for the interconnection of those types of networks.

The work that is described in the present paper is aimed to the performance study of two different approaches for LAN/WAN interconnection, that have been addressed by base standard activity (ISO, 1989; ISO, 1991a; ISO, 1991b), and that are currently being addressed by functional standardization efforts in ISO and in the regional workshops.

Section 2 introduces some basic concepts regarding the network layer and network layer relays, and presents the current status of functional standardization activities in the area. Section 3 presents a short description of the relay and traffic models used in the simulation studies. The study results are presented and analyzed in section 4. Special attention is given to throughput, transit delay, queue size and utilization data. Section 5 presents the study conclusions and future work directions.

2. NETWORK LAYER RELAYING

[1]European Workshop for Open Systems
[2]National Institute for Standards and Technology - OSI Implementors Workshop
[3]Asian and Oceanic Workshop

2.1 The OSI Network Layer

According to the principles of the OSI Basic Reference Model (ISO, 1984), the Network Layer provides the transparent transfer of data between transport entities, in such a way that the characteristics of different transmission and subnetwork technologies are masked and a consistent network service is offered.

In order to do so, the Network Layer is organized in three sublayers (Fig. 1) that may or may not be present in a system (ISO, 1988a), depending on the interconnected subnetworks:

- the SubNetwork Access Protocol (SNAcP) Sublayer, a subnetwork specific sublayer;
- the SubNetwork Independent Convergence Protocol (SNICP) sublayer, that presents an uniform service to the Transport Layer;
- the SubNetwork Dependent Convergence Protocol (SNDCP) sublayer, that is responsible for the necessary adaptations between the SNAcP and the SNICP sublayers.

In spite of this organization, the OSI environment supports two incompatible types of network services: the connection-mode network service (CONS), supported by the X.25 Packet Level Protocol (ISO, 1990), and the connectionless-mode network service (CLNS), supported by the connectionless-mode network protocol (ISO, 1988b). In the OSI environment, the interconnection of end-systems attached to the same or different subnetworks is only possible if the end-systems use the same type of network service. This CO/CL Interworking Problem has several possible solutions (EWOS,1990b; Schepers *et al.*, 1992a), all of them outside the context of OSI.

When the sub-network interconnection is carried out by network layer intermediate systems, those intermediate systems, or relays, must perform routing as well as switching functions. In addition, they harmonize differences between the interconnected sub-networks, and assure that the semantics of relayed information is preserved.

2.2 Types of relays

Depending on the way in which the information relaying is performed, relays may be grouped in two different types (EWOS, 1990b):

- Protocol relays, that relay the information on the basis of the semantics of protocol data units (PDUs) of a given layer, establishing a correspondence between the PDUs of one sub-network to the PDUs of other sub-networks (Fig. 2);

- Service relays, that relay the information on the basis of the semantics of the service supported by the protocols of the layer in which the relay operates. This approach requires the definition and use of an (N)-*Internal Layer Service* ((N)-ILS), that results from the addition of the necessary relaying functionality to the normal layer service (Fig. 3).

These types of relays can be used for the interconnection of different types of subnetworks, operating at one of several layers, and relaying one of the two service modes (connection-mode or connectionless-mode service). The corresponding relay identifiers are described in ISO (1991c), and are of the form

RXp.q

where

R = identifies the relay function

X = relay type identifier, covering the layer at which the relay operates, the service mode being supported and

the type of relay

p,q = subnetwork numerical identifiers

That is, RXp.q represents a relay of type X, between subnetwork type p and subnetwork type q.

2.3 Relay functional standardization

At the present moment, several functional standardization activities address network layer relaying, covering CLNS relaying (RAp.q profiles), CONS relaying (RBp.q profiles) and X.25 Packet Level Protocol relaying (RCp.q profiles). These functional specifications, or profiles, are being developed by EWOS, and are at different development stages (e.g., development in progress within organization, harmonization between regional workshops in progress, submitted to JTC1/SGFS for ISP processing). Current profile work addresses the interconnection of different types of subnetworks, e.g., CSMA/CD, Token Ring, PSDN, and FDDI, in various combinations.

Two examples of relay profiles regarding which there is a recognized interest are the profiles RB51.1111 and RC51.1111. The corresponding protocol stacks are shown in Fig. 4 and 5, respectively.

The RB51.1111 profile specifies a connection-mode network service relay for the interconnection of a CSMA/CD LAN (subnetwork identifier 51) and a PSDN. The access to the PSDN is permanent, via a PSTN leased line, and the relay supports virtual call establishment and termination procedures (subnetwork identifier 1111). The RC51.1111 profile specifies an X.25 protocol relay, for the interconnection of the above mentioned subnetwork types.

The performance study that is described in this paper was particularly aimed at these two profiles, but it is generally applicable to RB5n.1xxx and RC5n.1xxx relay profiles, that is, network layer service and protocol relays for the interconnection of local area networks and packet switched data networks.

3. STUDY DESCRIPTION

The relay study that is been carried out by the team addresses several aspects - not all of them covered by this paper - like, for instance, relay formal specifications, congestion control techniques, performance evaluation, management, and prototype implementations. In this paper we will address only the performance evaluation part.

In order to predict and evaluate the performance of network layer relays that use the service relay or protocol relay approaches, several simulation studies were carried out, based on different traffic conditions and relay models.

The use of simulation tools was chosen for a number of reasons:

- it enables great flexibility in changing traffic and relay characteristics, which can be used for a thorough study on the behavior of relays with different implementation options and under a variety of load conditions;
- it enables a study at different levels of detail, ranging from an analysis based on high level system building blocks, to the use of different machine processing powers and packet processing times;
- it can be used to obtain qualitative as well as quantitative data on the system performance;
- due to the computational complexity associated with the analysis of queueing

networks for the purpose of modelling and performance evaluation of computer and communication systems, the use of queueing theory is only feasible if several simplifying and restrictive assumptions are made (Schwartz, 1987), which only permits to obtain high level performance data, with reduced applicability;

- there aren't yet any standardized relay implementations that can be used for test purposes.

3.1 Relay models

The service and protocol relay models used in the simulation studies are represented in figures 6 and 7, respectively. The models take into account the effects of the physical and data link layers, that must be present in a real relay system (and that are, normally, taken care by dedicated communication controllers, with independent processing capabilities), as well as the effects of the particular building entities that exist in the network layer, that perform the relay functions described in ISO DIS 10028 (ISO, 1991a) for the service relay case, and ISO TR 10029 (ISO, 1989) for the protocol relay case.

The two subnetworks to which each of the relays is attached are, on one side, a CSMA/CD LAN and, on the other, a PSDN accessed via a permanent PSTN leased line at 128 kbps (see Fig. 8). Note that this 128 kbps leased line can represent the effect of several PSDN attachments at lower baud rates. Each of the relay models can deal with segmenting and reassembling operations.

For the purpose of the study, several operating facilities and parameters were modelled. This enabled the performance study to address different aspects of relay operation, and to highlight the behavior differences between service and protocol relays, under various relay configurations. For instance, the simulation models accommodated the use of several machine processing powers, line baud rates, and packet processing times. Window flow control mechanisms were not modelled, in order to extend the scope of the study to congestion collapse regions (Jain *et al.*, 1987).

3.2 Traffic conditions

The relays under study were submitted to various load patterns, representing a variety of network users and typical applications (e.g., electronic mail, file transfer, remote database access, virtual terminal).

Call generation obeyed a normal distribution which, due to the fact that the number of events per unit time was relatively high, is equivalent, with good approximation, to a Poisson distribution (Law and Kelton, 1982). The mean call interarrival time ranged from 20 to 300 ms.

During each virtual call the LAN module (Fig. 8) - that represents all the LAN users - sent a number of packets to the Relay module, ranging from 10 to 100, according to an uniform distribution.

In 40% of the calls, the data packets were small packets, representing traffic of interactive nature (e.g., virtual terminal applications, database queries). In 40% of the calls, the data packets were medium-size packets, that did not require segmenting or reassembly operations at the relay, and that represented e-mail or database access traffic. In 20% of the calls, the data packets were large packets, representing file transfer applications traffic.

3.3 Simulation strategies

The simulation studies were organized in three main stages. Table 1 summarizes the simulation conditions used in each phase.

The objective of the first stage was to determine the critical factors affecting the two types of relays under study and to obtain their general behavior characteristics. In this stage, the influence of the machine processing power - expressed in terms of executed operations per unit time - and the influence of the PSDN attachment line baud rate were studied. This permitted to choose the simulation parameters and conditions for the following simulation stages.

In the second stage the objective was to concentrate the analysis on the relaying capacity of the relay systems under study. In this stage the data transfer is unidireccional, with acknowledgements, but there are no call setup or clear call procedures.

The objective of the third stage was to study the global system LAN-RELAY-PSDN under different load and packet processing time conditions. In this stage all virtual call phases were modelled, i. e., call establishment, data transfer, and call termination.

4. STUDY RESULTS

4.1 First stage results

One of the factors that was studied in this stage was the influence of the PSDN attachment line baud rate. The simulation studies revealed that for baud rates up to 64 Kbps the service relay and protocol relay approaches are basically equivalent. This is due to the fact that the leased line baud rate constitutes a limiting factor for the amount of data that traverses the relay, which results in a very low relay utilization. Under these circumstances, the influence of the relay architectural choices is negligible, and thus the performance of the service relay and protocol relay is equivalent. At higher leased line baud rates the amount of data that passes through the relay becomes significant, and the relaying capacity appears as a determinant factor in the relay performance. In the subsequent simulation stages a 128 kbps baud rate line was used on the PSDN side. This line can represent the combined effect of several PSDN attachment lines, or an attachment line to a high speed PSDN (Sales, 1991).

The influence of the machine processing power was also addressed in this stage. Figures 9, 10 and 11 show some of the simulation results. In this stage the call interarrival time and the number of basic operations necessary to process a packet were fixed. The machine processing power varied from 1 million operations/second (ops/sec) to 40 million ops/sec.

In Figure 9 we can see that, with the used load and number operations to process a packet, the protocol relay and service relay utilization stabilizes for processing powers equal to or greater than 10 M ops/sec. This stabilization is imposed by the fixed transfer rate of the attached subnetworks and - in a smaller and decreasing factor - by the packet processing times. The difference between the service relay and protocol relay utilization remains constant, with a value of approximately 7%. In Figure 10 we can observe the corresponding effects in the maximum and average queue sizes. In this case the difference between the service relay and protocol relay queue sizes tends to decrease with the increase of the machine processing power. In Figure 11 we can compare the normalized throughput - defined as the ratio between the information that enters the relay and the information that leaves it - and the transit delay of small, medium and large packets for the service relay and protocol relay cases. Again, the stabilizing effect can be observed.

The above mentioned results point to one preliminary conclusion, that can be drawn out of this simulation stage: on one hand, for relatively low load conditions (i.e., low percent utilization), the protocol relay and service relay approaches can be looked at as equivalent in terms of overall performance; on the other hand, the two relay approaches behave quite differently when the load is increased, with important consequences in terms of queue sizes, throughput and transit delay.

In light of the simulation results of the first stage, the subsequent simulation stages were aimed to the study of the two relay types under heavy load conditions, in order to clearly characterize the behavior of the two types of relays, and to identify performance bottlenecks. In order to do this, all subsequent simulations used a simulated machine processing power of 1 M ops/sec and call interarrival times below and above 100 ms. This is equivalent to the use of higher machine processing powers and shorter call interarrival times. The relay percent utilization can be used for conversion purposes with some approximation, for processing powers up to one order of magnitude higher. The simulation results showed that for higher processing powers the numerical results cannot be taken directly from the 1 M ops/sec case, but the relative behavior of the relays remains basically the same.

4.2 Relaying capacity analysis

Figures 12 (curves RB-1 and RC-1), 13 and 14 show the simulation results of the second stage. In this stage the call interarrival times ranged from 300 ms to 50 ms, and the packet processing time was fixed and equal to 1 ms.

The graphics shown in these figures highlight the better relaying capacity of the protocol relay approach which, for a given load, presents less percent utilization, smaller queue sizes, higher throughput and smaller transit delay. Nevertheless, the difference in the performance of the service relay and protocol relay approaches becomes quite small for low traffic loads, as it was also apparent from the previous simulation stage,and remains with little significance for traffic loads corresponding to values as high as 70% for the service relay percent utilization.

The above mentioned results point to the conclusion that the service relay approach may be adequate for a relatively large number of cases. On the other hand, this simulation stage confirmed the expected superiority of the protocol relay approach to serve as a base for the implementation of network layer switching devices, when heavy load conditions are present.

4.3 Global system simulation results

The objective of the third simulation stage was to study the global system LAN-RELAY-PSDN under different load and packet processing time conditions, analyzing the combined effects of data transfer as well as call management procedures.

Figure 12 shows the first significant aspect that results from the consideration of call establishment and call termination procedures. When these procedures are modelled the obtained percent utilization is significantly smaller, which is due to the fact that data transfer procedures are conditioned by the completion of call establishment procedures. Thus, as is apparent from Fig. 12, the great difference in performance between the service relay and the protocol relay becomes much smaller when these relays are compared in an environment closer to the real environment. Another conclusion that can be drawn from these results is that service relays and protocol relays have equivalent behavior when the duration of the calls is relatively small, even for short call interarrival times. On the other hand, if the call duration is large and the data transfer is extensive the weight of the call management procedures becomes negligible and the relay utilization approaches that of curves RB-1 and RC-1, in Figure 12.

Another "real environment" effect can be observed in Figures 15 and 16. For call interarrival times bellow 50 ms, the protocol relay queue sizes and transit delays tend to get smaller. This could erroneously be interpreted as an increase in the protocol relay performance but, in fact, is due to an overload of Call Request packets. As the number of attempted call increases, the Relay and PSDN modules (see Fig. 8) are flooded with Call Request packets and only a small number of Call Accepted packets arrive to the LAN Module. This causes the number of Data packets generated by the LAN Module to decrease, and thus the relay queue sizes and transit delay also decrease.

As a final study in this simulation phase, four different sets of packet processing times were used. The obtained results are shown in Figures 17, 18 and 19, and are basically equivalent to those obtained in the second simulation phase. In fact, the increase of packet processing time can be seen as a load increase in terms of data and control packets.

5. CONCLUSION

The simulation study of the service relay and protocol relay philosophies for the interconnection of CSMA/CD LANs and PSDNs highlighted some aspects of relay performance, that may influence standardization activities, as well as implementation options and user selection decisions.

The studied relay types are currently being addressed by functional standardization efforts in ISO and in the regional workshops - particularly in EWOS - although at different development stages. The RCp.q relays are at a "submitted to JTC1/SGFS for ISP processing" stage, as long as the RBp.q relays are only at a "recognized interest" stage.

As a general conclusion, the presented study showed that both the service relay and the protocol relay types are valid and useful approaches to subnetworks interconnection. The choice should depend on a number of factors like, for instance, expected traffic loads, desired performance, available processing power, and implementation cost.

The study also revealed the sensitivity of the relay performance in relation to the architectural options involved. Processing entities, layer and sublayer interactions, and modularity are important and well established concepts, but can represent a handicap when high performance is required. Simulation has revealed itself as an excellent tool in the assessment of the architectural adequacy of a given relay approach.

The presented work has shown that the use of the RBp.q approach is adequate in the large number of cases where the traffic load is such that the relay utilization is not in excess of about 70%, requiring an amount of memory not much higher than that required by RCp.q relays, and having equivalent throughput and transit delay. This, and the fact that service relays have an architecture that derives from the OSI principles of layer interactions and structured approach, indicates that this type of relays requires a greater interest in what concerns functional standardization activities and product developers.

On the other hand, the simulation revealed the fact that RCp.q relays are necessary when heavy traffic loads are present and when there are requirements of high throughput, reduced transit delay and reduced relay memory. Thus, the study points to the conclusion that the protocol relay approach is more adequate than the service relay approach for high speed packet switching. This performance advantage is achieved at the cost of a greater implementation difficulty, that derives from the monolithic approach implied by the architecture of protocol relays, and may not be significant if the transfer rate of the attached subnetworks is not high, or is very dissimilar, or both.

The simulation study pointed out several topics and key areas that can be addressed in subsequent studies like, for instance, flow and congestion control capabilities, and interconnection of other types of subnetworks, with emphasis in the study of relay approaches for high speed network interconnection. These will be explored in parallel with the current work, that is aimed at the specification, study and development of prototype RB5n.1xxx and RC5n.1xxx relay implementations.

References

EWOS (1990a). 'OSI Layers 1 to 4 Addressing', *EWOS Technical Guide 002*, January 1990.

EWOS (1990b). 'Lower Layer Relays', *EWOS Technical Guide 006*, August 1990.

ISO (1984). 'Information Processing Systems - Open Systems Interconnection - Basic Reference Model', *ISO 7498*, October 1984.

ISO (1988a). 'Information Processing Systems - Data Communications - Internal Organization of the Network Layer', *ISO 8648*, February 1988.

ISO (1988b). 'Information Processing Systems - Data Communications - Protocol for Providing the Connectionless Mode Network Service', *ISO 8473*, December 1988.

ISO (1989). 'Operation of an X.25 Interworking Unit', *ISO Technical Report 10029*, March 1989.

ISO (1990). 'Information Technology - Data Communications - X.25 Packet Level Protocol for Data Terminal Equipment', *ISO/IEC 8208*, March 1990.

ISO (1991a). 'Information Processing Systems - Data Communications - Definition of the relaying functions of a Network Layer Intermediate System - Part 2: Connection-mode network service', *ISO/IEC DIS 10028-2*, 1991.

ISO (1991b). 'Information Technology - Data Communications - Intermediate system support of the OSI connection-mode network service using ISO/IEC 8208 in accordance with ISO/IEC 10028', *ISO/IEC DIS 10177*, 1991.

ISO (1991c). 'Information Technology - Framework and taxonomy of International Standardized Profiles - Part 2: Taxonomy of profiles', *ISO/IEC TR 10000-2*, 1991.

Law, A.M. and Kelton, D. (1982).*Simulation Modeling and Analysis*, McGraw-Hill Book Company, New York.

Jain, R., Ramakrishnan, K.K., Chi, D.M. (1987). 'Congestion Avoidance in Computer Networks With a Connectionless Network Layer', *DEC-TR-506*, Digital Equipment Corporation, August 1987.

Sales, B. (1991). 'A profile for wide area X.25 operating at 2 Mbps', *Computer Networks and ISDN Systems*, **21**, (4), 307-314.

Schepers, H.J.J.H., Rikkert de Koe, O.B.P., Havermans, G.M.J. and Hammer, D.K. (1992a). 'LAN/WAN interworking in the OSI environment', *Computer Networks and ISDN Systems*, **23**, (4), 253-266.

Schepers, H.J.J.H., Rikkert de Koe, O.B.P., Havermans, G.M.J. and Hammer, D.K. (1992b). 'Naming, addressing, routeing and relaying in the OSI environment', *Computer Networks and ISDN Systems*, **23**, (4), 241-251.

Schwartz, M. (1987).*Telecommunication Networks: Protocols, Modeling and Analysis*, Addison-Wesley Publishing Company, Reading, Mass.