NASA TECHNICAL REPORT

NASA TR R-475

NASA TR R-475 c.1

ADVANCES IN OPTIMAL ROUTING THROUGH COMPUTER NETWORKS

Israel M. Paz

Lyndon B. Johnson Space Center
Houston, Texas 77058

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • SEPTEMBER 1977

1. Report No. NASA TR R-475	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ADVANCES IN OPTIMAL ROUTING THROUGH COMPUTER NETWORKS		5. Report Date September 1977 6. Performing Organization Code JSC-09671
7. Author(s) Israel M. Paz		8. Performing Organization Report No. JSC S-446
9. Performing Organization Name and Address Lyndon B. Johnson Space Center Houston, Texas 77058		10. Work Unit No. 986-15-89-00-72 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Technical Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The optimal routing problem is defined. Progress in solving the problem during the previous decade is reviewed, with special emphasis on technical developments made during the last few years. The relationships between the routing, the throughput, and the switching technology used are discussed and their future trends are reviewed. Economic aspects are also briefly considered. Modern technical approaches for handling the routing problems and, more generally, the flow-control problems as presented at the Second International Conference on Computer Communications, held at Stockholm, Sweden, August 1974, are reviewed.

17. Key Words (Suggested by Author(s)) ·Switching Networks ·Store-and-Forward ·Circuit Switched Networks ·Flow Control ·Computer Communication Networks ·Optimal Routing	18. Distribution Statement STAR Subject Category: 66 (Systems Analysis)

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	19	$3.25

* For sale by the National Technical Information Service, Springfield, Virginia 22161

CONTENTS

ADVANCES IN OPTIMAL ROUTING

THROUGH COMPUTER NETWORKS

By Israel M. Paz*

Lyndon B. Johnson Space Center

SUMMARY

The routing-optimization problem is an important tool in the optimization of both the operational capabilities and the design of large-scale computer-communications networks. Extensive attention has been focused on routing optimization in the past 10 years because of the many large-scale computer networks that have been commissioned during this period of time. This report reviews different routing-optimization concepts that have been suggested, tested, and implemented in various parts of the world in the past few years. The complexity of the problem and the large number of theoretical and technical approaches suggested and tested have necessarily constrained this review to a limited group of specialists. However, many references are included to assist interested readers in locating technical details.

The main new approaches that characterize the technical advancements made recently in this field can be summarized as follows:

1. The development of new and more accurate models that not only calculate flows but also deal with the economic aspects of the optimal-routing problem

2. The differentiation between "through" and "local" traffic for alleviating congestions on a local basis

3. The use of "permits" and "permit budgets" for controlling flow at the node level or at the global level

4. The use of permits flow and other ingenious ways for updating the amount of permits allocated to a node or terminal

5. The idea of killing messages in overloaded networks on a global level and delaying or killing messages in congested areas or at the input terminals on a local basis, whenever these messages exceed a certain "age" in the network

6. The use of age for manipulating priority on a per-message level

7. The use of automatic path protection in packet-switching networks

---

*NRC-NASA visiting scientist.

A practical solution to the optimal-routing problem would probably use a combination of the proposed new techniques in a network that uses a combination of switching technologies to handle a certain combination of data flows of different characteristics. Additional work in the international standardization of hardware, software, and terminology is urgently needed.

INTRODUCTION

The routing-optimization problem primarily affects the operational aspects of computer network optimization. However, the choice of a routing policy also has far-reaching implications on optimal network design. Moreover, it has been shown that the same mathematical model could be used to handle both the optimal routing and the optimal redesign problems (ref. 1). Broadly stated, the problem can be formulated as the following flow-control optimization problem. Given a computer network specified by its topology, its link parameters (capacities, delays, costs, etc.), and its traffic demand statistics, a method is requested to control the flow of messages through the network in a way that optimizes a certain performance indicator while accommodating the maximum possible amount of traffic demand. Therefore, the routing problem is a network-control problem. The purpose of investigating the problem is to determine the best strategy for controlling flow through the network consistent with network specifications.

SWITCHING NETWORKS

The parameters suggested in the literature as performance indicators for the optimal-routing problem depend on the particular switching (flow control) technology used. A brief survey of the best known switching technologies and their implications on routing is therefore included. Detailed descriptions of these techniques are contained in references 2 to 4.

Store-and-Forward Switched Networks

One major type of switching networks is the store-and-forward (S/F) switched network. Included in the S/F type of network are message-switching (M.S.) and packet-switching (P.S.) networks.

Message-switching networks.- The main philosophy governing M.S. technology is that the network is requested first to accept (preferably all[1] (ref. 5)) incoming messages before any decision on how to handle and deliver them is made. This decision is determined only after the message has been accepted. Message switching therefore requires that the network should have a real-time input response to the external traffic demands; that is, an extremely large buffer memory.

[1]Recent techniques are relaxed on this requirement at heavy loads.

Another characteristic of M.S. technology is that the network retains control over the messages and the flow of messages through the network throughout the processing period. Only after the acknowledged error-free arrival and acceptance of a message at the destination or the initiation of a procedure to reject this message under overload conditions will this control be released.

Natural candidates for measuring performance in this type of network are the average or the maximum time delay (MTD) with which a message is delivered to its destination (refs. 6 and 7, respectively). An MTD is very sensitive to the traffic load on the network and to the manner in which the traffic load is distributed through the network (refs. 8 and 9, respectively). The MTD is continuously changing under the influence of variations in the traffic demand configurations and variations in the corresponding flow patterns induced in the network by a particular routing doctrine intended to cope with the variations in traffic demand configurations. Inadequate routing may allow long messages to produce congestion in the network flow. A result of such congestion is that the MTD will tend to exceed some practically acceptable values, which were recently used as thresholds for overflow control (e.g., MTD_{max} (ref. 8) or K_3 (ref. 10)). In such extreme situations, the throughput of the network will deteriorate and decrease to economically nonacceptable levels (refs. 2 and 8).

Packet-switching networks.- The P.S. technology introduced recently in computer networks evolved from the need to overcome the deficiencies of the M.S. approach (refs. 11 to 17). Although the S/F concept is preserved, a breakdown of each message into a number of packets of some standardized size is required in the P.S. technology. These packets travel through the network independently, not necessarily following the same path or the original time sequence, and are assembled at the receiving end by the network destination terminal.

Packet-switching technology enables the use of packing optimization techniques (some of which are reminiscent of equivalent techniques from the operational research field) to significantly increase throughput (ref. 18). Moreover, this improvement seems to require only a modest corresponding increase in the hardware complexity of the network. However, lack of standardization and the relative newness of the P.S. technology imposes an investment burden on the terminal equipment needed. This investment burden seems to have been one of the main reasons that discouraged the designers of the data network planned for West Germany from using the P.S. technology (ref. 19).

Successful implementation in the Advanced Research Projects Agency (ARPA) (ref. 12) and other computer networks (refs. 2, 20, and 21), adaptability to file-oriented processing, and proven superiority to the M.S. approach have made the P.S. network appealing. Therefore, designers of several computer networks that are in various stages of design and implementation have chosen to use the P.S. approach (refs. 13 to 15).

Circuit (Line) Switched Networks

The philosophy governing circuit (line) switched (L.S.) networks is to accept only those messages (traffic requests) for which a real-time communication channel can be established and to reject all others. The network therefore first searches for a path capable of serving each incoming traffic request (call). Only if such a path exists will the network assign it to serve this call, returning a busy signal otherwise. In this network, the user, after gaining access to the communication link assigned, has complete control over this link during the entire length of a call. This switching philosophy is as old as the telephone network, which is a classical example.

A good indicator of performance in L.S. networks should therefore be related to the number of calls that were rejected because the network could not secure a path for handling them. The most widely accepted performance indicator for this kind of switching is, as in telephone networks, the grade of service (G/S). Essentially, G/S is a measure of the percentage of calls requesting service that could not be handled by the network, averaged for a specific time. This measurement is clearly load sensitive. In computer networks, its definition will usually depart slightly from the definition that applies to telephone networks by using some kind of blocked-call-held approach. During the call-setup period, an L.S. network actually functions as an M.S. network when signaling information is concerned. A few data networks in various stages of design and implementation use this technology (refs. 19, 22, and 23). A detailed review of different computer networks (L.S. and S/F) is given by Allery in reference 24.

ROUTING, SWITCHING TECHNOLOGIES, AND FLOW CONTROL

Included in this section are discussions of routing and switching technologies; routing, switching, and performance; and first-generation routing techniques.

Routing and Switching Technologies

A detailed technical characterization of the three switching technologies previously described and full comparison of their relative merits is beyond the scope of this paper. The interested reader is referred to references 4, 16, and 25. Each of these technologies has its own merits for the particular application considered. A brief comparison of the relative merits of these technologies is presented in the following discussion.

The L.S. approach provides intrinsic code transparency, real-time (full duplex) connections, and immediate availability of an easily accessible and extremely reliable worldwide communication facility (i.e., telephone network), whereas the S/F approach usually provides better throughputs and therefore a higher operational efficiency. The S/F network is more naturally equipped to accept and interconnect, after proper translation, various communication codes of otherwise incompatible terminals. In fact, the S/F approach can easily provide more than a simple passive communication medium because flow control is handled by a computational facility under stored program control and because messages must be submitted in a properly formatted digital form. Thus, such functions as editing, prompting, storage for later retrieval, and priority handling on a per-message basis can be performed easily.

It is evident that the S/F approach integrates more naturally with computer technology. Moreover, designers with computer technology background can more easily predict S/F network performance because this prediction relies on the familiar theories of waiting lines and service scheduling. Therefore, it is not surprising that most of the computer networks designed during the period preceding the first International Conference on Computer Communications in 1972 (ICCC-72) and most of the designs that culminated from the search for efficient routing techniques embody the S/F approach in its P.S. version (refs. 4 and 25).

Some investigations into the theoretical (ref. 26) and practical (refs. 22, 27, and 28) aspects of optimal routing through L.S. networks have also been reported. However, the investigations did not have many practical implications until recently, when (as indicated subsequently) some of the characteristics of L.S. networks became economically appealing for certain applications (refs. 29 and 30).

The mixing of technologies in hybrid networks also has been attempted. The purpose was to use each technology in the range of its best performance. The AUTODIN network followed this path (refs. 31 and 32). However, the experience gained from operating the AUTODIN network is still controversial. Although the designers of AUTODIN appraise its performance highly, other leading computer network designers and analysts do not. In fact, some designers in an ICCC-74 panel discussion indicated, drawing from the experience gained with the AUTODIN network, that such a hybrid approach may inherit the disadvantages of both techniques and the benefits of neither.

At the ICCC-74 conference, it became clear that new and stronger consideration should be given to the merits of the L.S. approach, either as a solitary or an auxiliary technique. The main efforts in this direction are apparently led by Japanese researchers (refs. 29 and 30), who are motivated by throughput and efficiency of use considerations, and by West German researchers (ref. 19), who are motivated by short-range implementation considerations.

The previously mentioned classification of switching technologies has been popular for the past two decades. However, it provides only an ideal, abstract classification framework, and it should be used carefully. When practical networks are implemented, they usually require refinements that obscure classification boundaries. For example, not many S/F networks are actually designed to accept any input in real time. Some delay (ref. 5) (in modern designs (ref. 9), even loss) mechanisms are often introduced into the input/output ports to which the user terminals have access. Correspondingly, modern L.S. networks will delay, rather than reject altogether, some of the traffic submitted to their input ports during overload conditions (ref. 23). When handling the addressing information, they evidently use (as previously mentioned) the S/F approach during the path-selection and setup periods.

This classification of switching technologies into three general types has, however, some practical implications. Various routing techniques can more easily (sometimes only) be analyzed and compared when the switching technology that will implement them is specified. This classification is also a useful tool for analyzing the characteristics of the flow already in the network.

Routing, Switching, and Performance

Investigations into the relationships between throughput as the network performance indicator and the switching technology used, are described in references 28 and 29. Those investigations, especially the one by Itoh and others (ref. 29), may have triggered some of the new advances in routing technology that became public during the ICCC-74 conference. In these new routing techniques, the additional requirement for maximizing network throughput has been added to the more conventional G/S and delay considerations. The additional concern about throughput maximization is characteristic of what the author would like to consider as the second-generation routing techniques. Maximization is sometimes achieved at the expense of the G/S or MTD performances and may require the mixing of different switching technologies within the same system (refs. 3 and 4).

First-Generation Routing Techniques

The discussion of first-generation routing techniques includes the applicability of these early techniques in S/F networks and in L.S. networks.

In S/F networks.- A study of what the author would call first-generation routing techniques applicable to S/F computer networks is given by Fultz in reference 25. In this study, average message delay was used as the performance indicator to classify various routing techniques, to compare them theoretically and by simulation, and to suggest some adaptive routing algorithms for improving the performance indicator chosen. A concise but accurate review of the main terminology is given in reference 8. Those techniques seem to have culminated with the ARPA network that uses a one-parametric adaptive method called Shortest Queue plus Bias plus Periodic Updating (SQ+B+PUD).

In L.S. networks.- Studies of the first-generation routing techniques for L.S. networks are described in references 22, 23, 29, 30, and 33 to 35. The performance indicators are usually throughput or G/S. Some of the routing algorithms use shortest path techniques in weighted communication networks (ref. 35); other algorithms use fixed or adaptive routing tables (ref. 16).

ADVANCES IN OPTIMAL ROUTING

Accurate mathematical models for providing real-time solutions to the optimal routing problem under stochastic traffic demands have proved impractical (refs. 28 and 36). Most of the techniques proposed and implemented for solving the problem are therefore heuristic and are usually verified by extensive computer simulations (refs. 27, 33, 34, and 37). Only suboptimal results can be obtained by using these techniques.

In the period preceding the ICCC-72 conference, the emphasis in the development of such techniques has been on S/F networks of the P.S. type. Those efforts have culminated in the commissioning of a few experimental computer networks, the most representative of which are the ARPA network in the United States (ref. 12) and the National Physical Laboratories (NPL) network in the United Kingdom (ref. 21). The experience gained from these networks since their successful implementation has encouraged designers to propose and investigate the effects of introducing new refinements into their optimal-routing procedures and to suggest new techniques for improving performance. A recent review of various computer networks.in different stages of design and implementation is given by Allery in reference 24.

Refinements in Routing Techniques

The following discussion includes those refined and improved techniques that are considered to have led to a second generation of routing techniques. These advances are characterized by an increased concern with throughput maximization, by a tendency to mix switching technologies whenever and wherever possible, and by and increased interest in the possible implementation of these techniques in distributed networks (ref. 38). The purposes of refinements in optimal-routing technology can be described as follows.

Global level.- The main purposes of the refinements in routing techniques on a global level are to improve throughput for better plant use without excessive impairment of the G/S or the MTD performances of the network and to allow for technological implementations that will simplify network upgrading and expansion.

Local level.- The main purposes of the refinements on a local level are to facilitate their implementation on a node-by-node basis in distributed networks and to improve routing algorithms so that ping-ponging and looping (known in communication jargon as ring around the rosy situations) are minimized. Finally, changes in network and traffic conditions should be easily accommodated by adapting the routing policies at each node to the new situations, with minimal computational overhead and minimal increase in the flow-control-information traffic through the network.

Per-message level.- The main purpose of refinements on the per-message level is to enable easy trade-offs between throughput and delays under message control — for example, by aiming at short delays for interactive messages and high throughput for batch messages.

Main Tools for Improving Routing

The main tools that have recently been proposed and investigated for achieving the goals described in the previous section are discussed in the following paragraphs.

Global level.- The main tools for improving routing on the global level include mixing of switching technologies, input control, and standardization and modularization.

Mixing of switching technologies: After the relationship between the switching technology and the throughput has been established, approaches that have been proposed and investigated for mixing switching technologies to increase throughput can be used (refs. 3 and 4). In most systems, this mixing is implemented in a time division multiplexing loop (ref. 39) or switch (refs. 3 and 4). A fixed ratio between L.S. and S/F traffic (ref. 3) or a ratio that varies with the changing traffic demands could then be implemented to optimize performance (ref. 4).

Input control: To avoid a need for excessive buffer (memory) space in S/F networks or excessive trunk inefficiencies in L.S. networks (or both), special input control procedures have been proposed and implemented in modern networks (refs. 5 and 9). Input control is usually implemented by allowing for the destruction of messages that meet specified conditions in overloaded or locally congested S/F networks (ref. 9). Correspondingly, some delay in the input flow has been used in L.S. networks (refs. 5 and 23). In both situations, the calling party can be informed of the cause for the interruption or delay.

One way to provide such input control mechanisms is by the use of permits. A constant number (budget) of permits is allocated to the whole network in accordance with the network traffic-handling capability. These permits can then be used to control the input flow into the network by limiting the number of permits available at each considered input terminal.

Standardization and modularization: Modularization of hardware and software is being investigated constantly by international bodies and private companies. The investigation includes standardization of "handshaking" procedures, protocols, and interfaces (refs. 23, 24, 40, and 41).

Local level.- Routing improvement techniques at the local level include improvements in switching technology, improvements in routing procedures, and use of permits.

Improvements in switching technology: To improve switching technology, some novel time-division (multiplexing) switching systems are being investigated. A system proposed by the Centre National d'Etudes des Telecommunications (CNET) in France is designed for an L.S. service expected to be offered soon by the French Postes, Telegraphes et Telephones (PTT) (ref. 42). This system integrates switching and transmission in the final stage.

Improvements in routing procedures: A method for overcoming congestion at the local (switching node) level is to differentiate between through and local traffic and to give proper priority to the through traffic. Thus, separate buffer allocations for each type of traffic may be necessary.

Use of permits: The use of permits for regulating network flow by properly routing the flow is a rather new technique that originated from experimentation with the NPL (refs. 5 and 21) and the CYCLADES (ref. 13) computer networks. In networks using this technique, the number of information packets a node is permitted to handle is limited to the number of permits it is presently holding. Various flow-control and routing mechanisms are thus made possible. The mechanisms will vary with the manner in which the permits are assigned to a particular node and with the means by which the permits travel through the network. These permits may be allocated to a node in a rigid (ref. 21) or an adaptive way (refs. 5 and 13). The adaptive, or isarithmic, approach actually carries over to P.S. networks the idea of homogeneity in the flow-distribution scheme that was shown to help increase throughput in L.S. networks (ref. 14).

Another feature in this context is the inclusion of automatic route protection in P.S. networks. This feature, designed by Ferranti for the British Post Office, has been investigated in the Experimental P.S. Service (ref. 14). Two additional routing techniques in this investigation include using secondary routes if two attempts to use the primary routes fail, and never transmitting a packet through the branch in which it arrived at the node, even if this branch is part of a primary route. However, most of the routing, both in L.S. and S/F networks, still emphasizes look-up-table techniques, and most of the refinements only deal with the way those look-up tables are to be updated or used (e.g., by using the horizon concept (ref. 10)).

Per-message level.- Another recent technique is the use of aging parameters in routing. Aging parameters are aimed primarily at the reduction of ping-ponging and looping. In the time domain and on a per-message level, aging parameters perform a control action similar to the action performed by the permits on a local, per-node basis and in the traffic-volume domain. To illustrate, a node is temporarily "disabled" (saturated, congested) when all permits allotted to it have been exhausted. Correspondingly, a message is "killed" when it exceeds the maximum age $(MTD)_{max}$ it is permitted to acquire before being delivered (ref. 8). Aging parameters are also used in several ways for controlling traffic flow in the network. One method is to increase the priority of a message with its age. This last technique has been used with seemingly satisfactory results (refs. 8 and 10). The optimization of the threshold parameters representing maximum age (i.e., MTD_{max} (ref. 8) and the skip constant k_3 (ref. 10)) is discussed in reference 9.

Review of Recent Trends and Problems

Topics included in the review of recent trends and problems are mixing of switching technologies, synchronization problems, and economic considerations.

Mixing of switching technologies.- It has become evident that the P.S. technology (ref. 4) will not replace the L.S. technology in computer networks (refs. 40 and 41). Data networks will have to support both technologies (ref. 41). The relative performances of L.S. and S/F networks have recently been the subject of thorough investigations (refs. 2 and 29). In reference 29, properly normalized theoretical models are developed, and the results are tested on a practical network of typical characteristics. Some results of this investigation are summarized in the following paragraphs.

The traffic-handling capacities used in reference 29 to characterize a switching node or a transmission link are evaluated by the values of the call densities that they are capable of handling. These call densities are defined as the number of simultaneous calls that each of these elements are capable of processing under specified and practical conditions. For the switching nodes, those conditions are related to the processing time per call, consumed by the instruction set that handles each call. For the transmission links, the conditions are determined using their data signaling rates.

On the other hand, the cost indexes mentioned there represent a normalized measure of the installation cost per call. Therefore, they provide a unit of comparison for the cost effectiveness of the different networks considered.

The main conclusions obtained from reference 29 can now be summarized using the notations that follow.

Traffic-handling capacity: In an L.S. network, the traffic-handling capacity of a switching node exceeds that of a transmission link when long holding times are involved. For calls characterized by short holding times, this situation is reversed.

However, in S/F networks, the capacity of a transmission link almost always exceeds that of a switching node. Finally, the traffic-handling capacity of a transmission link is apparently always larger if used in a P.S. network than if used in an L.S. network.

Economies of growth: In S/F networks, both transmission and switching cost indexes increase with the information volume per call. Only transmission cost indexes do so in L.S. networks. Careful consideration should be given to this factor in the redesigning of a network intended to serve enlarged traffic demands, especially when the use of mixed switching technologies is contemplated.

Using the results of reference 29, a computer network designed to integrate L.S. and S/F switching technologies is proposed by Hirota, Kato, and Yoshida in reference 30. The network described in reference 30 is being investigated in Japan by the Nippon Telegraph and Telephone Company (NTT) in cooperation with four other private companies. Each terminal in this network is designed to enable the operator to select the type of switching technology that is required for a particular call by using a prespecified signaling sequence called discrimination.

Overhead: One of the important conclusions that can be drawn from reference 30 is that the effect of the additional overhead required by this hybrid switching approach on the delay times is negligible. This conclusion is reached by investigation into the number of processing steps needed per call (fig. 5.1 of ref. 30). Another important result obtained in reference 30 is that the number of processing steps per call is independent of either the data volume or the call duration times in L.S. environments. However, the number of steps increases almost linearly with those parameters in P.S. environments.

Optimal mix aspects: The optimal blend of traffic to be assigned to each of the two technologies is still being investigated. This problem (chiefly at the multiplexing level) has been investigated in Europe (refs. 3 and 4). Results indicate that a facility using both technologies and assigning traffic between them in a way that is adaptable to the randomly changing demand patterns could be designed to perform in an optimal way (ref. 4). The proposed solutions establish methods for providing L.S.-type channels through P.S. networks (ref. 3). Techniques used for this purpose include the complete reassembly of each message before delivery, the use of switching buffers for delay-jitter reduction, and the assignment of a higher priority to the L.S. type traffic, plus smoothing buffers where necessary.

Synchronization problems.- Synchronization problems are among the most challenging problems that occur from the mixing of switching technologies. A synchronization problem at the multiplexer level and a mutual synchronization problem at the global level arise in interconnecting computer networks (ref. 4). Because of this problem, designers strongly recommend synchronized switching in both L.S. and P.S. technologies (ref. 24).

Although L.S. networks are intended to carry traffic that is mostly asynchronously generated, the transmission of such traffic through the network could be either synchronous or asynchronous (ref. 2). A synchronous L.S. network is defined as a network through which the information flow is under clock control.

The bit synchronization that is needed is usually provided by the network when the network is clock controlled and by the terminal when the network is not clock controlled. To secure the necessary number of bit reversals and to facilitate bit-synchronization extraction, the data may require scrambling before transmission. In such cases, a means of bypassing the scrambling facility at the termination of a call must be provided.

Economic considerations.- The results of recent investigations have added the following main economic guidelines to computer network designs.

Network cost per terminal: The network cost per terminal has been shown to depend on the following factors. In L.S. technologies, this cost depends mainly on the traffic intensity per terminal A and on the average call duration time T. In S/F technologies this cost depends on A, T, and communication density (the ratio between the actual message transmission time and the overall call duration time).

Regions of most economical operation as related to the switching technology used could be approximated as follows (fig. 1.1 of ref. 30).[2] The L.S. technology is more economical than the others in the following regions. At low traffic, $A_1 < 0.05\,(1 - e^{-k_1 T})$ with $T > 10$ seconds and $k_1 > 0$. For calls of long durations, $0.2\,e^{-k_2 T} < A_2 < 0.5\,(1 - e^{-k_2 T})$ with $T > 180$ seconds and $k_2 > 0$; k_1 and k_2 are positive constants characteristic of the system. The S/F technology is more economical to apply for traffic intensities between the limiting values; for example, $A_1 < A_{S/F} < A_2$ for any T. Finally, digital leased lines seem to be most economical when large traffic intensities are involved; for example, $A > 0.5$ for any T.

It has been indicated that L.S. networks display a higher flexibility than other networks for growing economically from low (starting) traffic volume. The L.S. networks also integrate more easily than the others into existing Telex networks, and

[2]The bounds given here are obtained by rough approximations of the accurate results and are intended to provide a guideline only.

this aspect should be carefully considered when the existing Telex network is large. For example, the ease of integration has been a decisive factor in the design of the West German data network.

CONCLUDING REMARKS

The main trends that characterize the recent advances in optimal routing are summarized as follows.

Review of Ultimate Goals

Some changes have been made in the ultimate goals of computer communication network operation. More careful and more detailed investigations have been undertaken to accurately assess the interconnections between network throughput and the different parameters used to measure the network performance. Special attention has also been devoted to determine the economical implications of trading throughput for other performance measures.

Renewal of Interest in Mixing Switching Technologies

Renewed interest and investigations have resulted in progress in mixing switching technologies for more economical operation, more economical network expansion, and better handling of excessive congestion and delay situations.

Economic Considerations

From recent investigations, the following conclusions concerning costs in computer networks designs may be reached.

Network cost per terminal has been shown to depend on the following factors. In line-switching technologies, this cost depends on the traffic intensity per terminal and the average call duration time. In store-and-forward switching technologies, this cost depends on the traffic intensity per terminal, the call duration times, and the communication density (the ratio between the actual message transmission time and the overall call duration time).

Regions of most favorable operation have been evaluated for each switching technology and for digital leased lines. Considerations involving the economy of growth have indicated that flexibility for growing economically from low traffic volumes is best achieved in line-switched networks. These networks also integrate more easily with existing Telex networks.

Advancements in Routing Techniques

The recent advancements made in the optimal routing technology qualify it for second-generation status. Advances have been made on global, local, and per-message levels and in both hardware and software. A few of the new approaches that characterize the technical advancements made in this field are as follows.

1. The development of new and more accurate models for calculating flows and for dealing with the economic aspects of the optimal-routing problem

2. The differentiation between through and local traffic for alleviating congestions on a local basis

3. The use of permits for controlling flow, both at the global level and at the node

4. The use of permits-flow techniques and ingenious ways for updating the number of permits allocated to a node or terminal

5. The use of age to kill messages in overloaded networks on a global level or to delay or kill messages in congested areas or at the input terminals on a local basis

6. The use of age for manipulating priority on a per-message level (This technique is actually not a completely new idea, but in a new context, it has more powerful application possibilities.)

7. The use of routing tables, updated by sophisticated techniques

8. The use of automatic path protection in packet-switching using a limited and rather simple way of adaptive routing

9. A reiteration of the basic engineering understanding that a good solution will probably use a combination of the proposed new techniques in a network using a combination of switching technologies for handling a certain combination of data flows of different characteristics

The transition period in the emerging field of computer communications is not yet over; in fact, too many new techniques and ideas have already been generated. More solid work in the field of standardization of hardware, software, and terminology is now urgently needed in the international computer-communications community. Otherwise, interconnection of different networks will become an extremely difficult and sometimes even an impossible task, whenever the need for such interconnection will arise.

Lyndon B. Johnson Space Center
National Aeronautics and Space Administration
Houston, Texas, May 22, 1977
986-15-89-00-72

REFERENCES

1. Paz, Israel M.: Optimal Routing and Optimal Redesign in Communication Networks. Proceedings of the 7th National Convention of Electrical and Electronic Engineers in Israel. Apr. 1971, pp. 709-717.

2. Fick, Herbert: Structures and Operating Principles of Networks for Data Traffic. The Second International Conference on Computer Communication. Aug. 1974, pp. 525-533.

3. Kümmerle, Karl: Multiplexer Performance for Integrated Line- and Packet-Switched Traffic. The Second International Conference on Computer Communication. Aug. 1974, pp. 507-515.

4. Zafiropulo, Pitro: Flexible Multiplexing for Networks Supporting Line-Switched and Packet-Switched Data Traffic. The Second International Conference on Computer Communication. Aug. 1974, pp. 517-523.

5. Price, W. L.: Simulation Studies of an Isarithmically Controlled S/F Data Communication Network. Information Processing 74. Proceedings of IFIP Congress 74. North-Holland Publishing Company (Amsterdam), vol. 1, 1974, pp. 151-155.

6. Kleinrock, Leonard: Communication Nets, Stochastic Message Flow and Delay. McGraw-Hill Book Co., 1964.

7. Meister, Bernd; Müller, H. R.; and Rudin, Harry R., Jr.: New Optimization Criteria for Message-Switching Networks. IEEE Trans. Com., vol. COM-19, no. 3, 1971, pp. 256-260.

8. Pickholtz, Raymond L.; and McCoy, Caldwell, Jr.: Improvements in Routing in a Packet Switching Network. The Second International Conference on Computer Communication. Aug. 1974, pp. 249-251.

9. Jilek, Peter: Flow Control in Computer Networks. The Second International Conference on Computer Communication. Aug. 1974, pp. 239-247.

10. Cegrell, Torsten: A Routing Procedure for the TIDAS Message Switching Network. The Second International Conference on Computer Communication. Aug. 1974, pp. 253-262.

11. Davies, D. W.: The Principles of Data Communication Network for Computers and Remote Peripherals. Paper presented at IFIP Conference (Edinburgh, Scotland), Aug. 1968.

12. Wolf, E. C.: An Advanced Computer Communication Network. Paper presented at AIAA Computer Network System Conference (Huntsville, Alabama), Apr. 1973.

13. Pouzin, Louis: CIGALE the Packet Switching Machine of the CYCLADES Computer Network. Information Processing 74. Proceedings of the IFIP Congress 74. North-Holland Publishing Company (Amsterdam), vol. 1, 1974, pp. 155-157.

14. Pearson, David J.; and Wilken, Donald: Some Design Aspects of a Public Packet-Switched Network. The Second International Conference on Computer Communication. Aug. 1974, pp. 199-213.

15. Alarcía, Gabriel; and Herrera, Santiago: CTNE's Packet Switching Network. Its Applications. The Second International Conference on Computer Communication. Aug. 1974, pp. 163-170.

16. Després, Rémi: RCP, the Experimental Packet-Switched Data Transmission Service of the French PTT. The Second International Conference on Computer Communication. Aug. 1974, pp. 171-185.

17. Hayes, J. F.: Performance Models of an Experimental Computer Network. Bell System Technical Journal, vol. 53, no. 2, Feb. 1974, pp. 225-259.

18. Hillier, Frederick S.; and Lieberman, Gerald J.: Introduction to Operations Research. Holden-Day, Inc., 1967.

19. Elias, Dietrich: Application and Technical Aspects of Data Transmission in Western Europe and in the Federal Republic of Germany in Particular. The Second International Conference on Computer Communication. Aug. 1974, pp. 31-37.

20. Davies, Donald Watts; and Barber, Derek L. A.: Communication Networks for Computers. John Wiley and Sons, 1973.

21. Scantlebury, Roger A.; and Wilkenson, Peter T.: The National Physical Laboratory Data Communication Network. The Second International Conference on Computer Communication. Aug. 1974, pp. 223-228.

22. Gabler, Hermann G.: Plans for Data Communications in Germany. Proc. IEEE., vol. 60, no. 11, Nov. 1972, pp. 1374-1377.

23. Bothner-By, Halvor; Palonen, Vesa; Sjöström, Olov; and Svendsen, Hans: Study of Public Switched Synchronous Data Networks for the Nordic Countries. The Second International Conference on Computer Communication. Aug. 1974, pp. 271-278.

24. Allery, G. D.: Data Communications and Public Networks. Information Processing 74. Proceedings of IFIP Congress 74. North-Holland Publishing Company (Amsterdam), vol. 1, 1974, pp. 19-23.

25. Fultz, Gary Lee: Adaptive Routing Techniques for Message Switching Computer-Communication Networks. Ph. D. Dissertation, Univ. of California at Los Angeles, 1972.

26. Paz Israel M.: On the Traffic Handling Capabilities of Communication Networks Homogeneously Loaded. Networks, vol. 5, no. 2, 1975.

27. Paz, Israel M.; and Cederbaum, Israel: Evaluation of Optimal Policies in Communication Nets by Simulation. IEEE Trans. Com., vol. COM-20, no. 6, June 1972, pp. 264-274.

28. Paz, Israel M.: Routing Optimization in Communication Networks. D. Sc. Thesis, Senate of Technion - Israel Institute of Technology (Haifa, Israel), 1970.

29. Itoh, Kazuo; et al.: An Analysis of Traffic Handling Capacity of Packet Switched and Circuit Switched Networks. Proceedings of the Third Data Communications Symposium (St. Petersburg, Florida). Nov. 1973, pp. 29-37.

30. Hirota, Ken'Ichiro; Kato, Masao; and Yoshida, Yutaka: A Design of a Packet-Switching System. The Second International Conference on Computer Communication. Aug. 1974, pp. 151-162.

31. Jansson, H. A.: AUTODIN System Description - Part I Network and Subscriber Terminals. Western Union Technical Review, Jan. 1964, pp. 38-45.

32. Jansson, H. A.: AUTODIN System Description - Part II Circuit and Message Switching Centers. Western Union Technical Review, Apr. 1964, pp. 68-77.

33. Grandjean, Charles: Call Routing Strategies in Telecommunications Networks. Electrical Communications, vol. 42, no. 3, 1967, pp. 380-382.

34. Benes, V. F.: Programming and Control Problems Arising from Optimal Routing in Telephone Networks. Bell System Technical Journal, vol. 45, no. 9, Nov. 1966, pp. 1373-1438.

35. Cederbaum, Israel; and Paz, Israel M: On Optimal Routing Through Communication Nets. IEEE Trans. Com., vol. COM-21, no. 9, Aug. 1973, pp. 936-941.

36. Tomlin, J.A.: Minimum Cost Multicommodity Network Flows. Operations Research, vol. 44, no. 1, Jan. 1966, pp. 45-51.

37. Weber, J. H.: A Simulation Study of Routing and Control in Communication Networks. Bell System Technical Journal, vol. 43, Nov. 1964, pp. 2639-2676.

38. Baran, Paul: On Distributed Communication Networks. Trans IEEE Com. Syst., vol. CS-12, no. 1, Mar. 1964, pp. 1-9.

39. Sharma, R. L.; Shah, J. C.; El-Bardai, M. T.; and Sharma, K. K.: C-System: Multiprocessor Network Architecture. Information Processing 74. Proceedings of IFIP Congress 74. North-Holland Publishing Company (Amsterdam), vol. 1, 1974, pp. 19-23.

40. Doll, Dixon R.: Multiplexing and Concentration. Proc. IEEE, vol. 60, no. 11, Nov. 1972, pp. 1313-1321.

41. Nippon Telegraph and Telephone Corporation: Study and Development Planning of Digital Data Switching Systems. CCITT Contribution No. 75-E, Feb. 1974.

42. Texier, Alain: A Time Division Switch for Synchronous Classes. The Second International Conference on Computer Communication. Aug. 1974, pp. 265-269.

NASA-Langley, 1977 S-446

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

SPECIAL FOURTH-CLASS RATE
BOOK

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION
451

U.S.MAIL

600 001 C1 U G 770819 S00903DS
DEPT OF THE AIR FORCE
AF WEAPONS LABORATORY
ATTN: TECHNICAL LIBRARY (SUL)
KIRTLAND AFB NM 87117

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons. Also includes conference proceedings with either limited or unlimited distribution.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include final reports of major projects, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546