ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

Special Issue —
EIGHTH INTERNATIONAL
TELETRAFFIC CONGRESS
MELBOURNE, NOVEMBER 1976

Volume 11, Number 1, 1977

A.T.R. is issued normally twice a year (in May and November) by the Telecommunication Society of Australia. Each volume comprises two regular numbers issued in one calendar year.

Subscriptions for A.T.R. may be placed with the General Secretary, Telecommunication Society of Australia, Box 4050, G.P.O., Melbourne, Victoria, Australia, 3001. The subscription rates are detailed below. All rates are post free. Remittances should be made payable to the Telecommunications Society of Australia.

The Society and the Board of Editors are not reponsible for statements made or opinions expressed by authors of articles in this journal.

Contributions to A.T.R. should be addressed to: Secretary of Editorial Board, Australian Telecommunication Research, Telecom Australia Research Laboratories, 59 Little Collins Street, Melbourne, Australia 3000.

The Telecommunication Society of Australia publishes the following:

1. **The Telecommunication Journal of Australia** (3 issues per year)
   Subscription — Free to Members of the Society* resident in Australia
   Non-member in Australia $5
   Non-members or Members Overseas $7

2. **Australian Telecommunication Research** (2 issues per year)
   Subscription — To Members of the Society* resident in Australia $4
   Non-members in Australia $8
   Non-members or Members Overseas $10
   Single Copies — To Members of the Society resident in Australia $3
   Non-member within Australia $6
   Non-members or Members Overseas $7

* Membership of the Society $3

All overseas copies are sent post-free by surface mail.

Enquiries and Subscriptions for all publications may be addressed to:
The General Secretary, Telecommunication Society of Australia, Box 4050, G.P.O. Melbourne, Victoria, Australia, 3001.

ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

VOLUME 11, NUMBER 1, 1977

Contents

Challenge

Staging an International Teletraffic Congress in Australia must surely be the event of a lifetime - the Eighth ITC, held in Melbourne on 10-17 November 1976, was the first to be held in the Southern Hemisphere and only the second outside Europe. It presented the first opportunity for most Australian teletraffic engineers and research workers to participate in this international forum, where the world's top experts in this field could exchange experiences and ideas.

Australia started participating in these Congresses back in 1958, when it sent one delegate to ITC2 in the Hague to present a paper. Increasing contributions were made to subsequent ITC's, culminating in playing host to over 200 delegates from 27 countries at the Melbourne Congress.

The Editors of ATR decided to commemorate this important event by printing this special issue of the Journal. In addition to general information about the Eighth Congress and abstracts of all the papers presented, this issue also contains the papers by Australian authors, together with discussion. The papers are re-produced from original masters supplied by the authors. Other material has been printed in a uniform format and type size for consistency.

Eighth International Teletraffic Congress

J. RUBAS

Telecom Australia, Melbourne, Australia

The Eighth International Teletraffic Congress was held in Melbourne from 10 to 17 November, 1976. The Congress was held at the Southern Cross Hotel and was attended by 235 delegates from 27 countries. A total of 135 papers and reports were included in the technical programme, details of which appear in this issue together with abstracts of all papers. The papers by Australian authors are printed in full.

The first ITC was held at Copenhagen in June 1955, after a number of research workers in this field came to the conclusion that a meeting should be held to provide a forum for exchange of experiences and presentation of new ideas. The scope of the first congress is indicated by its full title : The First International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration. Copenhagen was chosen as the venue in recognition of the pioneering work done in traffic engineering by the Danish mathematician A.K. Erlang.

The first congress was a resounding success and it was decided to hold similar gatherings in the future. An International Advisory Council (IAC) was set up to control and guide the organisation of future congresses. The detailed organisation and staging of a particular congress would be the responsibility of a national organising committee formed in the chosen host country, which would also bear all costs. The current membership of the International Advisory Council is given in the Appendix.

All subsequent ITC's were held within this organisational framework, at three year intervals: the second at the Hague in 1958, the third in Paris (1961), the fourth in London (1964), the fifth in New York (1967), the sixth in Munich (1970), the seventh in Stockholm (1973), and the eighth in Melbourne. The next ITC will be held in Spain in 1979.

The size and scope of ITC's grew with passing years. The first ITC was attended by 69 delegates from 13 countries, with 26 papers presented and discussed in four working days. By the seventh ITC the number of delegates had grown to 328, the number of countries represented to 30, and the number of papers to 130, which required six working days to present and discuss. Because of remoteness from other population centres, the number of delegates at the Melbourne Congress was lower than at the 7th ITC in Stockholm but the number of papers presented (135) set a new record.

The technical programme of ITC 8 is given on subsequent pages of this special issue. It was arranged into 26 working sessions (not counting the opening and closing ceremonies). Following the practice introduced at the New York Congress (ITC 5), the programme includes a number of invited papers on selected topics. These papers, written by acknowledged experts in their fields, are intended to review progress to date and to inform on the current state of the art in various areas of traffic engineering. Eight invited papers were presented at the Melbourne Congress; they are marked in the programme by an asterisk. Two of the invited papers were tributes to outstanding members of the traffic engineering community who died since the previous congress - E.P.G. Wright and Dr. Yngve Rapp. Other invited papers dealt with traffic engineering topics of current interest.

The quality of the papers presented and the level of discussion at the Melbourne Congress maintained the high standards established in the previous congresses. It is not intended here to review the papers presented, since their contents are adequately described by the abstracts. As indicated by the abstracts, all areas of traffic engineering and some related subjects have been covered. Dimensioning methods, analysis of delay systems, S.P.C. switching systems, and traffic data measurement and administration attracted most attention and required two sessions each to deal with all the papers covering these subjects. To permit oral presentation of at least half the papers offered and to leave adequate time for discussion at the end of each session, parallel sessions had to be held during one afternoon.

The staging of an ITC in Australia presented the first opportunity for many Australian traffic engineers and research workers to participate in this international forum. For most overseas delegates it meant meeting old friends, making new acquaintances and visiting a new continent. Many new ideas were presented and discussed during the congress and added to the growing body of knowledge about teletraffic engineering. The record of discussions in each technical session will be forwarded for inclusion in the proceedings of the congress, which were distributed to all delegates. The full proceedings will also be supplied to selected libraries in various overseas countries. In Australia, a full set of papers presented at this and all previous ITC's is held at the Telecom Australia's Headquarters Engineering Library in Melbourne.

REFERENCES

1. The First International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration. "Teleteknik", 1 (1957) : 1, pp. 2-130, English edition.
2. The Second International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration. "Het PTT-bedrijf", 9 (1960) : 4, pp. 158-209.
3. Troisième Congrès International de Télétrafic, Paris, 11-16 Septembre 1961. "Annales des Télécommunications". 17 (1962) : 7-8 pp. 145-226. Special issue. In French and English.
4. Report of the Proceedings of the Fourth International Teletraffic Congress, London 1964. "Post Office Telecommunications Journal". London 1965. Special edition, pp. 1-68.
5. Preprints of Technical Papers, Fifth International Teletraffic Congress. New York 1967.
6. Record and Discussions, Fifth International Teletraffic Congress, Vol. II. New York 1968.
7. Sixth International Teletraffic Congress. Munich 1970.
8. Proceedings of the Seventh International Teletraffic Congress. Stockholm 1973.
9. Jacobaeus, C. : ITC - The International Teletraffic Congresses. "Ericsson Review", Vol. 50 (1973), No. 4.
10. Olsson, K.M., Anderberg, M. and Lind, G. Report on the Seventh International Teletraffic Congress in Stockholm, June 13-20, 1973. "Ericsson Technics", Vol. 29 (1973), No. 3.

APPENDIX

LIST OF COUNCIL AND COMMITTEE MEMBERS

International Advisory Council

Prof. - Dr. A. Jensen, Denmark - Chairman
Mr. W.S. Hayward, U.S.A.
Dr. C. Jacobaeus, Sweden
Prof. - Dr. L. Kosten, Netherlands
Dr. P. Le Gall, France
Mr. J.A. Povey, U.K.
Dr. C.W. Pratt, Australia
Dr. K. Rohde, F.R. of Germany
Mr. I. Tånge, Sweden
Mr. R.I. Wilkinson, U.S.A. - Honorary Member

At the end of the Melbourne Congress, Mr. Tånge retired from the IAC and Prof. Kosten was made a honorary member. At the same time, four new members were elected to the Council : Prof. J.W. Cohen (Netherlands), Dr. G. Gosztony (Hungary), Dr. V. Neiman (U.S.S.R.) and Mr. J. Villar (Spain).

National Organising Committee

Mr. I.A. Newstead, Australian Telecommunications Commission - Chairman
Mr. P.S. Bethell, Australian Telecommunications Commission
Mr. J.A. Black, Australian Telecommunications Commission
Mr. T. Broughton, Australian Telecommunications Commission
Mr. K.S. Brown, Standard Telephones and Cables Pty. Ltd. (Sydney) Australia
Mr. S. Cronstedt, L.M. Ericsson Pty. Ltd., Australia
Prof. R.B. Potts, University of Adelaide
Dr. C.W. Pratt, Australian Telecommunications Commission
Mr. J. Rubas, Australian Telecommunications Commission
Mr. J.M. Ryan, Australian Telecommunications Commission

Technical Programme Committee

Dr. C.W. Pratt, Australia - Chairman
Dr. L.T.M. Berry, Australia
Mr. W.D. Eccleston, Australia
Mr. A.H. Freeman, Australia
Mr. P.J. Lenn, Australia
Mr. J. Rubas, Australia
Mr. R.C. Walker, Australia

Prof. J.W. Cohen, Netherlands
Prof. H. Inose, Japan
Prof. - Dr. A. Lotze, F.R. of Germany
Dr. K.M. Olsson, Sweden
Mr. J.E. Villar, Spain
Dr. E. Wolman, U.S.A.

Technical Programme

* Invited Paper.

ANALYSIS OF DELAY SYSTEMS I

Chairman : *Cohen, J.W. (Netherlands)*
Discussion Leader : *Segal, M. (U.S.A.)*

TRAFFIC DATA MEASUREMENT AND ADMINISTRATION 1

Chairman : *Hayward, W.S. (U.S.A.)*
Discussion Leader : *Lee, L. (Canada)*

ANALYSIS OF DELAY SYSTEMS II

Chairman : *Cohen, J.W. (Netherlands)*
Discussion Leader : *Smith, J.L. (U.K.)*

TRAFFIC DATA MEASUREMENT AND ADMINISTRATION II

Chairman : *Hayward, W.S. (U.S.A.)*
Discussion Leader : *Iversen, V.B. (Denmark)*

NEW METHODS OF ANALYSIS

Chairman : *Wallström, B. (Sweden)*
Discussion Leader : *Pearce, C. (Australia)*

SUBSCRIBER BEHAVIOUR PROBLEMS

Chairman : *Wilkinson, R.I. (U.S.A.)*
Discussion Leader : *Barnes, D.H. (U.S.A.)*

* Invited Paper.

* Invited Paper.

NETWORK AND SYSTEM OPTIMISATION

Chairman : *Neiman, V.I. (U.S.S.R.)*
Discussion Leader : *Anderberg, M. (Sweden)*

DATA TRAFFIC PROBLEMS

Chairman : *Elldin, A. (Sweden)*
Discussion Leader : *Wolman, E. (U.S.A.)*

INTERNATIONAL AND I.T.U. STUDIES

Chairman : *Hofstetter, H. (Germany)*
Discussion Leader : *Newstead, I.A. (Australia)*

LINK SYSTEMS

Chairman : *Jacobaeus, Chr. (Sweden)*
Discussion Leader : *de Boer, J. (Netherlands)*

S.P.C. SWITCHING SYSTEMS I

Chairman : *Inose, H. (Japan)*

S.P.C. SWITCHING SYSTEMS II

Chairman : *Inose, H. (Japan)*
Discussion Leaders : *Dietrich, G. (Germany)*
Gerrand, P. (Australia)

* Invited Paper.

STATISTICAL ANALYSIS OF TRAFFIC DATA

Chairman : *Rahko, K. (Finland)*
Discussion Leader : *Olsson, K.M. (Sweden)*

REPORTS

Chairman : *Rubas, J. (Australia)*

FINAL PANEL DISCUSSION

Chairman : *Jensen, A. (Denmark)*

CLOSING CEREMONY

Chairman : *Jensen, A. (Denmark)*

Abstracts of the Papers

ANALYSIS OF BUSY SIGNAL SYSTEMS

"E.P.G." - A Tribute to E.P.G. Wright. 121
Gimpelson, L.A. and Rice, J.

To summarise a career which spanned 50 years and a multiplicity of disciplines related to telephony is a difficult task. To couple that with an insight into the enthusiastic, contributing, influential, competitive and colorful personality behind that career is impossible. Those who knew and worked with EPG should excuse the brevity of this tribute, which will touch on EPG's traffic work only.

On the Conditional Blocking Probabilities in a Link System for Preselection. 122
Michaux, J.

In modern telephone exchanges conditional selection is commonly used to control the multistage switching networks. The loss computations then become very difficult because of the existence of internal blocking. A rigorous solution cannot be obtained for link systems of practical size.

In this paper the faithfulness of two approximate calculation methods has been evaluated with respect to two sample preselection networks. It is shown that the methods assuming statistical independence only are more faithful than the methods assuming functional and statistical independence. A significant trend towards too high computed c.b.p's is however detected in one stage of the network. An attempt is therefore made to improve the faithfulness of the calculation method through a partial release of the assumption of statistical independence. It is shown that the calculated c.b.p's then become much more accurate and, as a consequence of the equations, so do the p.d's and the value of B. It is worth noting that the computational complexity is not increased by the suggested modification of the calculation method.

Functional Algorithm and Structural Switching of Special Automatic Telephone Information Equipment. 123
Dimchev, M.I., Halachev, V.I. and Ivanov, P.D.

In this present paper we consider the results of the planning of the special automatic inquiry information equipment (ANIE) for q x 18 telephone service of a city automatic telephone exchange (ATZ) on the basis of the methodology (arrived at by the authors) of the logical synthesis of the block structure of the automated information equipment for public use (AIE-ÖA). Basic requirements of the controlling system are analysed, on which basis the functional algorithm has been put together for the controlling equipment and transmission adaptation equipment, which serve for the connection of the equipment to the ATZ.

The construction has been based on the specially designed equipment for 18-channel tape recordings.

The aim of the work is not only to provide the results of the design of a concrete ANIE, but also the laying down of several methodical basic directions in the design of the AIE-ÖA with varied application, principally using the methods of communications technology on one of the least elaborated stages of the planning, the block synthesis.

It is proposed that the final objective of the block synthesis of the AIE-ÖA be the assembling of the block functional diagram directly from the minimum form of the function algorithm.

In this paper we have included the block functional diagrams of the general structure (Fig. 4) and of the transmission adaptation equipment (ÜAE) (Fig. 5), and also the basic diagram of the relay variant (perfected by the authors) of the ÜAE for use with the ATZ of the electro-mechanical type, for example A-29.

On Erlang's Formula for the Loss System M/G/K. 124
Cohen, J.W.

It is a well known fact that Erlang's famous loss formula for the blocking probability of a full availability group of K lines with Poissonian input only depends, in so far as it concerns the holding time distribution, on its first moment and not on the particular form of this distribution.

This phenomenon has drawn the attention of many investigators and quite a number of proofs of it are available in literature. Nearly all of these proofs are not very acceptable from a mathematical point of view, and if they are acceptable they are rather intricate, and do not give an insight in the real facts behind this interesting phenomenon.

In the present study a proof of Erlang's formula is given which is simple and avoids intricate computations, in particular the use of differential equations. On the other hand it leans heavily on a number of fundamental theorems from the theory of regenerative processes and Markov processes. As a byproduct of the investigation a simple proof is obtained of the fact that in the stationary situation the departure process is a Poisson process if blocked arrivals are also considered as departures.

General Telecommunications Traffic Without Delay. 125
Le Gall, P.

It is intended to demonstrate how the general formulae relating to telephone traffic processes without waiting may be obtained. The lost call and repeat attempt models are considered, essentially for the case of an arbitrary service time distribution. The explicit formulae are then extremely complicated when the arrival process is general. They become simple only for a certain category of processes which are called here "pseudo-poissonian". This simplicity takes the form of a simple distribution function for the "remaining service times" of the "indistinguishable" calls in progress, a distribution already well known for the Erlang model.

Two cases are then examined for which the preceding results may no longer be valid because of the non-independence of the arrival process in relation to the epochs of system congestion : firstly, the case of overflow traffic and then that of repeat attempts. The consequences of the dependence are not the same in the two cases.

The Influence of a Preceding Selector Stage on the Loss of Gradings. 126
Bazlen, D. and Lörcher, W.

This paper deals with investigations on the influence of a preceding selector stage on the loss of gradings. Several models of preceding selector stages are considered which differ with respect to the size of the preceding stage, the wiring mode between the outlets of the preceding stage and the inlets of the grading and the hunting mode of these outlets, resp. These investigations were performed by extensive simulation. Recommendations are given for practical applications how to interconnect consecutive selector stages in step-by-step systems.

DIMENSIONING METHODS I

Structural Properties of Connecting Networks. 131
Neiman, V.I.

This paper discusses the structures of space-division connecting networks, for which crosspoint minimisation problems are solved. Theoretical studies of bounds for minimum crosspoint number are reviewed. Discussion of connecting networks designs focuses on the structures of non-blocking simultaneous connecting networks (re-arrangeable systems). Other classes of blocking and non-blocking networks are briefly described.

Multihour Engineering of Alternate-Route Networks. 132
Eisenberg, M.

This paper describes a procedure, called "multihour engineering", which is used to engineer traffic networks for more than one hour of point-to-point load data. We describe the results of a study of multihour engineering in a large-scale local network, as well as an experiement which was undertaken to determine whether predicted multihour savings could be realised in an actual network. Some of the practical aspects of implementing multihour engineering are discussed.

A Model Relating Measurement and Forecast Errors to the Provisioning of Direct Final Trunk Groups. 133
Franks, R.L., Heffes, H., Holtzman, J.M. and Horing, S.

This paper describes a mathematical model of the provisioning of direct final trunk groups with forecasting and measurement errors. This model can be used to study the effects of applying standard trunking formulae to possibly inaccurate load forecasts.

Using the mathematical model, a set of curves known as the Trunk Provisioning Operating Characteristics is calculated. These relate percentage of reserve capacity to service (as measured by the fraction of trunk groups with blocking exceeding 0.03). The accuracy of the estimate of the traffic load defines the curve on which one is constrained to operate. The degree of reluctance to remove trunks together with the traffic growth rate determines the operating point. Improved estimation accuracy corresponds to a more desirable operating characteristic.

The accuracy of the forecasted load estimate is influenced by many factors. Data base errors (e.g., measuring the wrong quantity due to wiring or other problems), recording errors (e.g., key punch errors) and projection ratio errors illustrate some of these factors. This type of modeling may be useful both in evaluating the potential effects of proposed improvements in measurement or forecasting accuracy, and in studying the effects of changes in provisioning policy.

Traffic Models for the Traffic Service System (TSS). 134
Augustus, J.H.

This paper describes the traffic models developed for dimensioning operator positions and the switching network in the Traffic Service System (TSS). TSS is a stored program common control system designed to process operator assisted traffic for small sized toll applications. The models have been validated by simulation and a comparison of theory and simulation results is included in the paper.

Equivalent High Usage Circuits and Their Application in the Analysis and Design of Local Networks Employing Alternate Routing. 135
Harrington, J.S.

The cost of routing traffic between origin and destination exchanges employing alternate routing can be expressed in terms of the direct route alone by adding to it a number of circuits to equal the cost of routing the traffic, overflowing from the direct route, over the alternate routes of the network. EQUIVALENT HIGH USAGE CIRCUITS, the sum of the actual and additional direct route circuits, are shown to be functions of offered traffic, marginal occupancy (sometimes referred to as cost factor) and efficiency of the traffic switching machine. Tables of equivalent high usage circuits, for a range of marginal occupancy values and covering both full and limited availability trunking with various link losses, are presented. Examples are given in the use of tables for optimisation of availability, comparisons of switching equipment and local network analysis. Curve fitting equations, relating actual and equivalent high usage circuits to pure chance offered traffic with marginal occupancy as parameter, are given for full availability trunking. (This paper is reproduced in full on pp. 29-37)

An Engineering Method to Account for Link Congestion Effects on Dial Tone Delay. 136
Guess, H.A. and Kappel, J.G.

A new mathematical model for estimating dial tone delay in No. 5 Crossbar Switching machines is reviewed. This model demonstrates that a small number of lines, waiting for dial tone on concentrators that have all of their links busy, can generate a disproportionately large number of ineffective marker seizures. This causes worse than expected dial tone delay to all subscribers in the office.

Comparisons of results from the model with field data show generally fair agreement. For engineering purposes the high day service threshold is adjusted downward to account for the expected variability of high day load conditions. An engineering procedure based on results from the analytical model is presented and illustrated.

DIMENSIONING METHODS II

Some Practical Problems of the Traffic Engineering of Overloaded Telephone Networks. 141
Gosztony, G. and Honi, G.

In overloaded networks it is difficult to determine the basic data of traffic engineering. Some measurements were made to assess the interrelation between the traffic situation and the relative weights of the reasons of failure. A mathematical model of approximate nature allows to decide whether a traditional or a "repeated attempts" model should be adopted in dimensioning parts of a connection path. The data required for this are only the carried traffic, the number of call attempts and the holding times. The mathematical model presented for "repeated attempts" purposes assumes that the investigated group has in spite of repetitions an Erlang or Engset character with some fictitious offered traffic. Other parts of the networks and the called subscribers are also considered. This model neglects reattempt times but can take several failure type dependent perserverance functions into account. Simulation tests have shown that calculation results may be used in a wide range of practical cases.

A Method for the Calculation of Traffics Offered to an Alternative Routing Arrangement. 142
Sagerholm, B.

In planning multi-exchange networks the traffic distribution in the network must be calculated for several years ahead. For this calculation we need the existing traffic distribution in the network.

In the Swedish network all routes are measured two times each year. These measurements will give carried traffic per route in erlang. One important task is to calculate the part of the total offered traffic to each route which constitutes direct offered traffic, as the offered traffics to the routes are a mixture of both direct offered traffic and overflow traffic from other routes.

The method in this report determines the different offered traffics to an alternative routing arrangement in such a way that the sum of squared relative differences between calculated and observed carried traffic per route is minimised.

Carried traffics are calculated using the Wilkinsons equivalent random method. The method is extended to accept offered smooth traffics. The calculations are performed using a standard iterative routine for minimising a sum of squares. The initial values of the offered traffics are obtained by a "backwards" application of the Wilkinsons method.

A computer program has been written for all these calculations and has been tested on several different routing arrangements with satisfying result.

Dimensioning of Traffic Routes According to the EERT-Method and Corresponding Methods. 143
Rahko, K.

In this paper the ERT, EERT, NDM and WDM dimensioning methods are compared. It is proved that the NDM and (E)ERT are valid dimensioning methods based on different criteria The NDM is based on the time congestion criteria. For EERT it is calculated values according to time congestion (EERT-T) and compared both methods in this case. Methods based on time congestion improve the dimensioning of telephone networks on the basis of overall grade of service because then the total congestion can be determined as a sum of the time congestion in successive routes. Whether the NDM is better than the EERT can be proved only by measurements.

Dimensioning of Alternative Routing Networks Offered Smooth Traffic. 144
Rubas, J.

This paper reviews the methods of dimensioning networks employing alternative routing and discusses the use of Binomial distribution model when the offered traffic is smoother than pure chance. Instead of the Poisson-based "equivalent random" method the more accurate direct computation of overflow traffic moments is proposed. A comparison is made between this and two other methods for accuracy and convenience of computation. Sample dimensioning graphs for full and limited availability trunk groups are appended.
(This paper is printed in full on pp. 38-44)

Computations with Smooth Traffics and the Wormald Chart. 145
Nightingale, D.T.

The paper presents an evaluation of the mathematical expressions for overflow traffic when the number of circuits assumes real quantities. Efficient computational methods are given for evaluation of overflow traffic to an arbitrary accuracy.

The resulting methods are applied to Wilkinson's equivalent random theory (ERT-W) for both rough and smooth traffics. A new ERT(ERT-N) is derived and applied which avoids the iterative complications of ERT-W. An equivalent binomial theory for rough and smooth traffics is defined and results compared with the ERT's and exact results. An equivalent random queue theory is defined for several queue disciplines and a brief comparison made with exact results. A directly computed Wormald chart and a table of values for $E(-1,A)$ and $\psi(0,A)$ are included where $\psi(0,A) = (\partial \ln \Gamma(N+1,A)/\partial N)_{N=-1} - \ln A$.
(This paper is printed in full on pp. 45-52)

Modular Engineering of Junction Groups in Metropolitan Telephone Networks. 146
Farr, J.P.

If a telephone network is designed so that each junction route is dimensioned to the nearest circuit, then when the network is re-dimensioned some time later to cater for changes in point-to-point traffics the new design would probably show that most of the routes should be changed by at least one circuit. On the other hand, if the network is designed so that each route is dimensioned to the nearest preferred modular size (e.g. 5, 10, 15, 20, etc.), then when the network is re-dimensioned some time later for changed traffics, it is likely that only a small proportion of routes would need to be changed. This paper gives a comparison of circuit requirements and costs for a real metropolitan network configured with different choices of module size. The paper also includes comparative statistics for different choices of module size on the number of routes which would require a change in size at the time the network goes through successive stages of re-dimensioning to cater for the change in traffic with time. A strong case is made for the adoption of modular engineering as a design principle for metropolitan networks employing alternative routing.
(This paper is printed in full on pp. 53-59)

FORECASTING METHODS

Forecasting Telephone Traffic in the Bell System. 211
Blair, N.D.

This paper describes the methods presently used by the Bell System for forecasting telephone traffic. Summaries of the forecasting techniques used for Central Office switching systems, interoffice trunks, and operator forces are given. A wide variety of methods are used ranging from simple extrapolation to complex mathematical models. Whatever method is used, it must be tempered with informed judgement based upon an understanding of the fundamental characteristics of the business and careful analysis of all significant information.

An Econometric Forecasting Model of the Demand for International Voice Telecommunication from Canada. 212
Khadem, R.

The purpose of this study is twofold. Firstly, to focus research on the determinants of demand for long distance communications to overseas countries, to single out those variables found to be significant in explaining demand, and to estimate the functional relationships; and secondly, to suggest a new approach to describing the demand process. Previous researchers have analysed the demand process by using a single equation to express the relationship between the quantity demanded and its causes. An approach entailing the estimation of a three-equation system offers a far better approximation of the demand process. This approach is referred to as the "access-usage" approach.

Traffic Forecasting with an Inadequate but Improving Data Base. 213
Turner, W.M. and Willett, R.B.

This paper deals with the production of forecasts of busy hour traffic flows for exchange planning purposes. The nature of traffic series is frequently such that established forecasting methods are inappropriate. The paper reviews the problems faced by British Post Office staff at the local level and describes the methods being developed to overcome them. The basic principle is one of computer produced time trend extrapolation of traffic quantities and calling rates followed by reconciliation of discrepancies between related forecasts, to produce estimates of levels up to 7 years ahead.

Use of Computer Programs for Trunk Forecasting. 214
Gitten, L.J.

The efficient design and operation of a network requires the co-ordinated efforts of a great many people. In order to gain high network efficiency they should be able to take advantage of sophisticated traffic theories, analyse and manage large quantities of data, and co-ordinate their individual efforts. To accomplish this for the more than 300,000 trunk groups which make up the Bell System message network, it has been necessary to develop a package of computer programs as part of a Total Network Data System.

The four programs making up this package are the Common Update System which maintains the records which describe the network, the Traffic Data Administration System which provides for storage, summary and screening of traffic data, the Trunk Servicing System which develops the base year traffic loads and the Trunk Forecasting System which calculates growth ratios and provides estimates of future trunk requirements. This paper describes these systems, with emphasis on their capabilities and their interfaces with the people who must operate them. These programs have currently been installed in more than half of the Bell System's administrations with further conversions scheduled.

NETWORK PLANNING AND DESIGN

Yngve Rapp - A Memoir. 221
Jacobaeus, Chr.

This paper briefly surveys the work of the late Dr. Yngve Rapp who died on 12 March, 1976, at the age of 72. His work dealt almost exclusively with the planning, dimensioning and construction of telephone networks. He has made many important contributions to this field, which are briefly reviewed in this paper.

A Mathematical Model for the Long Term Planning of a Telephone Network. 222
Bruyn, S.

This paper presents a mathematical model for the long term planning of a telephone junction network. The use of this model to minimise the cost of a network whilst maintaining grade of service requirements at each time period results in a large non-linear programming problem. A dynamic programming algorithm to solve this problem is presented and applied to a practical problem.
(This paper is printed in full on pp. 60-64)

Computer Aided Planning of the Telephone Network of Rural Areas. 223
Caballero, P.A., Sole Gil, J., Knutsen, K.M. and Hansen, I.

A new computer program for the economic optimisation of rural telephone networks and its application to a real life planning study is presented. The paper first analyses the experience gained with previous programs as the starting point for the specification of the new program. Next, the main characteristics of the new program are described, including its input/output data and logic. The first application of the program is a planning study in Norway, which has a double purpose: to make the planning study itself and to evaluate the applicability of the program. Both subjects are analysed in the last sections of the paper.

Hypothesis of a Toll Network with Separate Routes for First-Choice and Overflow Traffic. 224
Diano, G., Pallotta, P. and Masetti, G.

The paper develops the subject, dealt with previously, of finding a hierarchical toll structure which would not only be the economic optimum but would also operate well in overload conditions.

The structure envisaged, called book-network because of the configuration of its routing chart, has the characteristic of always keeping-choice and overflow traffic separate by splitting up the trunk groups and the transit exchanges.

Having defined the calculation methodology and established the design of this new network, it is compared with the traditional one from the point of view of costs and of efficiency, both in design and overload conditions; the latter are taken, for the sake of simplicity, as shortage of circuits on high-usage or final trunk groups.

The comparison shows that the book-network is more capable of guaranteeing, in overload conditions, the handling of the traffic of the small relations but is more expensive than the traditional network; however, when it becomes necessary to split up transit exchanges in operation the difference in cost becomes negligible.

A Nodal Route Switching Network Composed of Standard Modules. 225
Grimbly, J.L.C. and Smith, J.L.

This paper presents a form of nodal switchblock which enables the blocking characteristics and pattern of availability of a small switchblock to be extended to a wide range of larger switchblocks. This principle can be applied so that growth can be achieved without reconfiguring previously established links and terminations. The traffic characteristics of typical implementations are examined and suitable applications for the switchblock in communications networks are suggested.

Computer Aided Study for the Determination of the Structure of the Algerian Long Distance Network. 227
Belhadj, A. and Caballero, P.A.

To define the long term structure of the Algerian long distance network is the aim of the computer aided study presented here. This paper deals with the methods used to prepare the input information, the main emphasis being placed on the traffic forecast procedures used. Traffic sensitivity analysis and comparative studies of alternative solutions are the major points discussed. Conclusions are given for this particular study. Then, general qualitative conclusions, applying to this class of problems, are drawn from the study. The paper ends with a short description of the optimisation program used, including a discussion of its major algorithm.

ANALYSIS OF DELAY SYSTEMS 1

A Queueing System with Time-Outs and Random Departures. 231
Wallström, B.

This paper presents a study of an M/M/n queueing system whose customers may leave the queue or the server group before service completion. These departures are governed by the following assumptions:

(i) A customer's waiting time in the queue is limited by a constant as well as by a random variable with exponential distribution.

(ii) The service given to a customer is limited by another random variable with exponential distribution.

The model was developed as a tool for analysing certain switching devices in telephone plant such as registers and markers, under heavy loading conditions. Thus the constant limitation of the waiting time may mocel the effect of automatic time outs while the random limitations of the waiting and service times may correspond to the events, when subscribers decide - for one reason or another - to hang up before connection is established.

An extension of the model assumes individual constant waiting time limitations for different customer streams. This may be applied, e.g., to the queues in front of register groups in transit exchanges, where different time out settings can be used to give some traffics higher priority than others.

Queueing Model with Regular Service Interruptions. 232
Fujiki, M. and Murao, Y.

This paper is dovoted to the study of delay which is caused by regular service interruptions. In order to evaluate delay in call processing programs in the electronic switching system, a queueing model with many priority classes is formulated. The significant assumptions of the model are that : 1) The highest priority class is a pre-emptive priority class and arrives at a single service station at discrete and uniform levels. 2) The second and lower priority classes arrive at the station in groups at uniform intervals, identical to those of the highest class. 3) The service times of the highest priority class follow a distribution which is derived by the convolution of the unit distribution and the negative exponential distribution. The service times of the lower priority classes follow negative exponential distributions. The means and the distributions of the waiting and completion times for the second and lower priority classes are derived. The results by calculation are shown to be in good agreement with those by simulation.

Approximations for Certain Congestion Functions in Single Server Queueing Systems. 233
Delbrouck, L.E.N.

A simple way to approximate the delay distribution in a M/G/1 system with first-come first-served (FCFS) or random order of service (SIRO) queue discipline is to take as approximation thereof the simplest distribution that agrees with it in terms of two parameters - e.g. - the first two moments - and admits the same heavy traffic approximation.

On the other hand, M/G/1 busy period distributions are too heterogeneous to admit such a simple minded treatment. However, their residual counterparts share a number of interesting properties with M/M/1 busy period distributions and conditions are discussed under which the latter may serve as two-parameter approximations for the former. We also discuss the representation of the M/M/1 busy period distribution as a probability mixture of Erlang distributions of odd integer order which is particularly well suited for computational work.

With regard to the FCFS delay distributions of M/G/1 systems, particularly when service time distributions are discrete with finite support, the original simple approximation may often be improved upon in a manner that brings into play the influence of the service time distribution without affecting the two-parameter match.

In conclusion, we discuss somewhat more involved applications of the two-parameter match in the estimation of FCFS delay distributions in GI/G/1 systems.

Queue Discipline NIFO for a Tree Structured Generalised M/G/1-Queueing System and Its Application (NIFO = Nearest In First Out). 234
Egenolf, F.J.

Certain tree structured switching networks can be treated as a generalised M/G/1 queue, where the service discipline is characterised by the term "Nearest In First Out" (NIFO). In this queueing system the service times are mutually dependent - a fact which aggravates the mathemtical treatment of the model.

After discussing in detail the behaviour of the NIFO-model several variants of the model are developed which differ in their tree structure and also in their "secondary" discipline (FIFO or LIFO) supplementing the "primary" discipline NIFO. The distributions and mean values of the delay time, the service time and the number of calls in the system are obtained by extensive simulation. As a direct application the special multiplexor of a data switching system is described.

Approximate Formulae for the Delay in the Queueing System GI/G/1. 235
Kramer, W. and Lagenbach-Belz, M.

The general single server system GI/G/1 has been treated manifold, but only for some special cases handy formulae are available. Very often exact calculations are too cumbersome and sophisticated for practical engineering, as well as upper and lower bounds generally are too gross approximations.

Therefore the need was felt to support the traffic engineer with simple explicit approximation formulae, based on a 2-moments approximation.

In this paper such formulae are derived heuristically for the mean wating time and the probability of waiting.

The quality of the formulae, which have been checked by numerous comparisons with exact and simulation results, is such, that within the most interesting range of server utilisations from 0.2 to 0.9 the error is less than 20% (typically $<$ 10%) for all combinations of the arrival and service processes characterised by the following distribution types : D, E_4, E_2, M, H_2.

By known relations, also simple approximations are provided, e.g. for the variances of the associated output processes, the first two moments of the idle time distribution and the mean length of a busy period.

Analysis of Complex Queueing Networks by Decomposition. 236
Kühn, P.

In this paper an approximate method for the analysis of complex queueing networks is proposed. The queueing network is of the open network type having N single server queueing stations with arbitrary interconnections. There is only one class of customers (calls) which arrive acc. to general exogenous arrival processes. The service times of the queueing stations are generally distributed. The analysis is based on the method of decomposition, where the total network is broken up into subsystems, e.g., queueing stations of the type G/G/1. The subsystems are analysed individually by assuming renewal arrival and departure processes. All related processes are considered with respect to their first two moments only. An analysis procedure is reported which reduces the total problem to a number of elementary operations which can be performed very quickly with the aid of a computer. Numerical results are reported to demonstrate the accuracy of the method. The paper concludes with a discussion on extensions of the method.

Provision of Signalling Equipment According to a Delay Criterion. 237
Johnsen, S. and Smith, J.L.

When a signalling system such as R2 allows a delay between the request for signalling equipment, and that equipment actually becoming available to handle a telephone call, it would appear reasonable to adopt a delay criterion for deciding the quantity of equipment to be provided. That this is not generally done is due to a number of reasons such as the absence of an agreed network performance specification in terms of delay, the difficulty of performing the calculations, and uncertainty whether it would on balance be beneficial to design a network in this way. This paper proposes a number of performance criteria and examines the consequences of applying them in a certain hypothetical network, and against the background of conventional electromechanical switching equipment.

TRAFFIC DATA MEASUREMENT AND ADMINISTRATION 1

Traffic Data - The Need, Nature and Use. 241
O'Shaughnessy, J.J.

This paper on traffic data :

1) Explores the nature of the traffic function in an Operating Telephone Company and demonstrates its dependency on traffic data.

2) Reports on two studies carried out in Bell Canada dealing with :

   (a) The sensitivity of capital expenditures to the accuracy of traffic data and forecasts.

   (b) The use of time series and cross-sectional analyses to produce more accurate switching centre and administrative area forecasts of traffic usage.

3) Emphasises the need - with the availability of more and better traffic data - to re-evaluate current service criteria and to develop new or revised traffic theoretical approaches.

Extreme Value Engineering of Small Switching Offices. 242
Barnes, D.H.

This paper describes a series of traffic studies and a resulting plan to implement for small switching offices mechanised dial administration and traffic engineering using observations of extreme traffic values. The work is an extension of that discussed by E. Wolman at the 7th ITC(11) which concerned the application of extreme-value distribution theory to the provisioning of line concentrators. Small switching offices of 2000 lines or less comprise more than half of Bell System buildings and there are many times more customer switching systems (PBXs). These offices cannot economically support conventional measurement systems, but still a way to collect and process data for accomplishing better service and cost control is needed. The use of peak values fills this need by greatly reducing both the equipment required and the amount of data to be collected and processed. It also makes possible engineering service criteria which better reflect the customer's experience in using the system.

The complexity involved requires the use of a computer. Described is a fully mechanised data system, a simple design of pollable central office hardware and a mini-computer central control which can communicate with a very large number of locations over a single data port by polling each office just once a day or once a week during low traffic periods. The computer has communication programs, an operating system, calculation algorithms and report formats which should furnish in a fully mechanised manner the administrative and engineering information needed for small offices.

Cost Effectiveness of Traffic Measurements. 243
Chin, Y.M.

This paper is concerned with the optimisation of the method, where the data collected from a previous measurement plays an important part in the measurement-dimensioning cycle.

As any traffic measurement is only an estimate of the offered traffic, further usage of the data will result in errors, directly caused by the sampling process. If the sampling variance is known, this imprecision can be allowed for in the dimensioning process. This adjustment represents an additional cost penalty attributed to the practicality of traffic measurements.

In this paper, it is assumed that the cost of the traffic study is linearly related to the duration of the measurement. The cost penalty is shown to be approximately inversely proportional to the time spent in conducting the experiment. If we consider the measurement phase as part of an investment, clearly, by choosing the traffic measuring parameters, the return on investment can be maximised.

The above concept is applied to the occupancy measurement of a first choice high usage route in a simple triangular alternative routing pattern. The traditional network cost minimisation technique and full availability working are assumed. (This paper is printed in full on pp. 65-70)

Traffic Measurements and the Grade of Service. 244
Enke, T. and Rahko, K.

A lot of traffic measurements have been made in Finland using various methods. The results of these measurements are reported on the basis of the busy-hour, busy-period and 24-hour day concept. The correlation between measured traffic intensity and the grade of service is discussed.

Several traffic measuring devices have been developed. In this report, an automatic measuring device is described in detail. The device enables the measuring of traffic intensity simultaneously on several routes. Traffic measurement results are collected in a memory. Any desired measuring period may be selected, and the output from the memory can be directly in the form of mean values for desired traffic periods.

Accuracy Requirements Concerning Routine Traffic Measurements, with Regard to Service Level Objectives in Telephone Network and to Certain Error and Cost Factors. 245
Parviala, A.

The task of routine traffic measurements is firstly to check if the traffic needs are satisfied at the moment, and secondly to give a basis for estimations concerning the future development of the traffic and for plans to increase the number of lines in order to satisfy future demands, too.

This presentation is a study of the error components inevitably occurring in traffic measurements, in basic data for prognosis and in planned operations, in relation to the real need for lines, defined by the grade of service aimed at, but not visible until later.

With examples based on statistics and special practical studies it is shown which error components, if reduced, primarily have the most significant effect economically.

This examination gives some guidelines as to the choice of measuring methods sufficiently qualified for different cases.

ANALYSIS OF DELAY SYSTEMS II

Graded Delay Systems with Infinite or Finite Source Traffic and Exponential or Constant Holding Time. 251
Kampe, G. and Kühn, P.

The paper deals with single-stage delay systems of the types M/M/n and M/D/n and single-stage delay-loss systems of the type M/M/n-s both in case of full or limited accessibility. Further distinctions are made between single and multi-queues, ideal and real gradings, and a finite or infinite number of traffic sources. Waiting calls are served acc. to the disciplines FIFO, RANDOM or LIFO.

Starting from known results, various systems are reviewed systematically with respect to their stationary state probabilities, characteristic mean values and waiting time distributions. Curves of calculated results are given and compared with simulations.

Waiting Time Distribution in Link Systems. 252
Haugen, R.B. and Østerud, H.

This paper deals with approximate calculation methods for delay link systems with several input queues. The calculations are based on the idea of "equivalent systems" where the link system is equated with :

(i) a full availability group, and

(ii) Erlang's ideal grading.

An equivalent system also includes an equivalent queueing structure which might be completely different from the original one.

The equivalent system is constructed to have the same probability of waiting, W, as the link system. In the latter case, W might be calculated from the modified methods of Lotze or Jacobaeus, but the possibility of improving the approximation for W is also suggested. This improvement leads to a better approximation of the mean waiting time τ than found by Hieber. Finally, an equivalent queueing structure for the ideal grading is constructed and approximate formulas for the waiting time distribution are found.

A Network Flow Model Analysis of a Multi-Queue Operator Service System with Priority. 253
Liu, J. and Kettler, D.A.

A multiqueue operator service system with server priority is formulated as a bipartite congestive network flow problem. The intensity of the intraflow on each arc is dependent on the congestive conditions at each node. Feedback equations are introduced which relate the congestive conditions at each node to the intensity of the intraflow on each arc. This network flow model is transformed to a fixed point problem. It is further shown that if the congestive functions associated with each node of the network are continuous with respect to load, then a fixed point always exists. For specific load and server parameters, the fixed point is interpreted as the flow intensity on each arc under an equilibrium state.

This mathematical model is applied to the traffic characterisation of a large call distributor system handling directory assistance calls. This particular distributor allows a limited number of calls which are blocked at the preferred server group to intraflow to nonpreferred groups with an idle server. A preferred group of servers primarily handles calls from a geographical cluster of customers. Thus the preferred server has a shorter serving time than a nonpreferred server. Yet because of the intraflow capability, the efficiencies of large team operation are retained. Empirical data support the intraflow traffic characteristics predicted by the mathematical model.

Through the use of this model the amount of intraflow for each cluster to nonpreferred servers becomes a predictable quantity for a forecasted offered load and specified server parameters. Thus, results generated by the model become an integral component in the determination of the minimum number of operators in each group required to provide objective service to the customer.

Accuracy of Observed Mean Values of Waiting Times, Queue Lengths and Proportion of Delayed Calls in Markovian Many-Server Systems. 254
Olsson, K.M.

By using formulas from renewal theory asymptotic approximations are derived for the variance of the observed mean waiting time, mean queue length and proportion of delayed calls during a given time interval in some markovian service systems. Numerical results are given for the M/M/c waiting systems.

Turfing - A Queueing Model for the Work Backlog of a Telephone Repairman/Installer. 255
Segal, M.

Recent studies have demonstrated an increase in the productivity and motivation of a worker when he could claim a "thing of his own". A trial is now underway in which each telephone repairman/installer is assigned a well-defined geographical area - a "turf". This paper describes a model of the work backlog in a turf and presents some results from the trial.

A Calculation Method for Link Systems with Delay. 256
Villen, M.

In this paper an analytical method of calculating the waiting time distribution in a multi-stage link system is presented.

The paper treats a multi-stage link system with conditional selection. The arrival of calls is assumed to be of either the Poisson or Engset-Bernoulli type. The calls finding congestion wait in a queue, for which several alternative disciplines can be defined.

The calculation method is applied to two practical cases with 3 and 4 switching stages respectively and the numerical results obtained are compared with simulation results.

TRAFFIC DATA MEASUREMENT AND ADMINISTRATION II

The Total Network Data System. 261
Buchner, M.M. and Hayward, W.S.

The increasing complexity of telephone networks and the need for immediate information for network management have increased the need for complete, processed data. No longer can engineering judgement satisfactorily compensate for the omissions and errors that occur when large volumes of traffic data are collected and processed manually.

This paper presents the concept of a network data system and describes the Total Network Data System (TNDS) now being constructed and implemented in the Bell System. TNDS provides the capability for handling network data from the source in the switching machine to the ultimate user such as the network manager, network administrator, or network engineer. New machines have been designed for the acquisition and rapid dissemination of data; a major part of the development effort has been in the area of general purpose computer programs which transform raw data into meaningful, validated network information. By the end of 1976 the Bell System will be more than halfway to complete implementation of TNDS.

Busy Hour Traffic Variations Determined from Continuous Traffic Measurements. 262
Leigh, R.B. and Little, A.J.

Since January 1973 the UK Post Office has been using an experimental traffic recorder at one of its medium sized trunk exchanges which records the number of circuits engaged minute by minute throughout the day directly onto magnetic tape, thus producing actual information on traffic flow. The paper describes the measurements taken, indicates the aims and area of study, summarises the results obtained, and enumerates any conclusions reached.

The Rigorous Calculation of the Blocking Probability and Its Application in Traffic Measurement. 263
Mina, R.R.

The first part of this paper deals with the theory for the rigorous calculation of the mean and variance of the load offered to a group of fully available trunks and its blocking probability from the mean and variance of the carried load.

The second part describes a concept for an inexpensive traffic measuring system. The system is processor controlled and makes use of the computing power of the processor in calculating the mean and variance of the offered load and its blocking probability from the mean and variance of the load carried by a limited portion of the trunks which are arranged to be used when all the other trunks in the group are blocked or occupied. The system also includes facility for event count measurement and for visual display of congestion in overloaded groups as well as an indication of plant overprovision in underloaded groups.

A Study of Subscriber Dialled Trunk Traffic in the Short Term. 264
Cole, A.C. and MacFadyen, N.W.

We present some preliminary results of a long-term investigation into the day by day behaviour of telephone traffic in the U.K. In particular, we introduce a new method for describing and parametrising the all-day traffic profile.

Customer Line Usage Studies. 265
Hartman, M.G.

Due to the impact that proposed rate structures and technological advances are likely to have on telephone traffic in the future, GTE has undertaken two customer line usage studies in its domestic telephone operating companies. In these studies, information is recorded on each call made on individual studied lines. This information is combined with customer data in downstream processing to produce a detailed call data base.

In this paper, the reasons that these two studies were undertaken are outlined, the study plans and hardware are described, the results of some preliminary analyses are presented and discussed, and future studies on the data are described. In the results section, the distribution of call arrivals and the distribution of message holding times are analysed. In addition, some usage statistics by class of service are derived and applied to an analysis of load balancing techniques.

An Improvement in Traffic Matrices Calculation. 266
Marin Martin, J.I.

A statistical proceding for the treatment of the experimental data is described, consistent in the application of the method of maximum probability in order to obtain the best values. Otherwise, this method can be easily calculated with the help of a computer.

NEW METHODS OF ANALYSIS

Operational Research Methods in Traffic Engineering. 311
Fujiki, M.

The purpose of this paper is to stress the importance of operational research techniques in the solution of problems encountered in teletraffic theory. Methods now in use in the operational research field are too numerous to permit their enumeration here, and therefore, this survey is confined to optimisation problems.

In this review, some general optimisation techniques are considered, together with areas in which these methods have been successfully applied. For continuous variables, various gradient based methods have been used to determine minimum cost configurations in multistage alternate routing networks. Penalty function methods (e.g. the SUMT method) have been successfully applied to the optimal design of computer communication networks. For discrete variables, integer programming has been used in the design of multidrop line networks connecting remote terminals to a central data processing centre. The principles of dynamic programming have been incorporated into models involving multistage decision processes.

The paper concludes with some comments on optimal control problems and identifies areas which may benefit from application of optimising techniques.

On General Point Processes in Teletraffic Theory with Applications to Measurements and Simulation. 312
Iversen, V.B.

Earlier works on this subject deal mainly with Markovian traffic processes, and they are based on the state probabilities of the system considered. This paper assumes the arrival process and the holding times are stochastically independent, but place no restrictions on the traffic process. A few basic elements of the theory of point processes are mentioned.

The applied measuring principle is either a continuous measuring method or a scanning method. We first consider the measuring of a single time interval in an unlimited and a limited measuring period respectively, and we achieve many new results. Then we analyse the traffic volume or intensity by adding the calls occurring within the measuring period. Previous results on the observed traffic load are derived in an easy way, and furthermore, the theory includes arbitrary holding times, arbitrary scan intervals and intensity variations.

Applications of Processing State Transition Diagrams to Traffic Engineering. 313
Gerrand, P.

The likely adoption by the VIth Plenary Assembly in October 1976 of the graphical Specification and Description Language (SDL) prepared by the CCITT's Study Group XI offers potential advantages to teletraffic engineers. This paper introduces the SDL, and suggests a general scheme for the systematic application of the SDL to system documentation, whereby the documentation required for capacity studies can be generated as a natural part of the system design process. The usefulness of processing state transition diagrams in general, of which the SDL is a special but important case, to both simulation and analysis of traffic capacity is discussed.
(This paper is printed in full on pp. 71-83)

Decomposition Techniques for Evaluating Network Reliability. 314
Fratta, L. and Montanari, U.

In this paper an efficient technique to evaluate the terminal reliability of a network consisting of unreliable independent undirected arcs is presented. This technique is an extension of the quite useful series-parallel reductions to the case where it is possible to isolate subnetworks of the given network connected to the rest of the network through three or more nodes. It is shown that this technique, based on recursive decompositions, leads to a linear computational complexity for any class of "n-m structured" networks.

Uncertainty Model for an M/M/1 System. 315
Mizuki, M.

The author examined in his ITC7 paper the effect of relaxation of probability axioms, whereby the additivity would not hold.

An examination of uncertainty problems shows that a realisation, which is in strict sense fixed before and after its observation, must satisfy the logical constraints of inclusion and dichotomy. Certain sub-structures of a Boolean lattice, called filters and intervals, are compatible with these constraints, i.e., sets belonging to such classes satisfy both inclusion and dichotomy. A probability measure must always be defined with respect to a Boolean algebra. When its domain of definition is reduced to filters and intervals then the resulting restriction of a probability measure no longer satisfies the probability axioms, and its additivity is replaced by partial ordering.

The simple M/M/1 queueing model problem is examined using the uncertainty theoretic approach as the main topic of this paper. Instead of using the balancing equation of birth and death process model, the queue size sequences are directly analysed. This elementary approach is consistent with the theory and yields results which are in agreement with observations.

Some Applications to Telephone Traffic Theory Based on Functional Limit Laws for Cumulative Processes. 316
Lindberger, K.

Cumulative processes appear frequently in traffic theory. If regeneration points in time can be chosen so, that the intervals between those points are i.i.d. r.v.'s. and the behaviour of the increments in the process over these intervals also are i.i.d., then we can call the process cumulative. Let e.g. the beginning of each congestion period be a regeneration point, then the total time congestion in (0, t] can be studied as a cumulative process, w(t).

Some results can be achieved from functional limit theorems for w(t).

To apply those theorems some constants have to be known. In simple cases they can be calculated, but in more complicated situations simulation is more useful. To illustrate the special technique and the advantages of having the results in a functional form, we have chosen some simple processes, where the norming constants are known. Functionals, stopping times and especially the random change of time method will be discussed. Applications to scanning processes are also given.

Average Number of Disjoint Available Paths. 317
Timperi, G.

The computation of probabilities associated with combinations of events may be a laborious task due to a combinatorial complexity. This kind of difficulty can often be reduced by resorting to Boolean algebra and graph theory. The computation of the average number of disjoint available paths is considered in the paper as a particularly illustrative example.

SUBSCRIBER BEHAVIOUR PROBLEMS

Comparison of Calculated and Simulated Results for Trunk Groups with Repeated Attempts. 321
Gosztony, G.

Call attempts arriving to a fully available trunk group could encounter the failures: congestion, no answer, busy called party. The parameters as call set up times, perseverance functions, distributions of reattempt intervals were related to the types of failure and were independent of their relative weight. All parameters were derived or originated from measurements. Simulation with average perseverances and reattempt intervals underestimates the harmful effects of repetitions. The averages themselves vary as traffic and failure situations are varied this should also be taken into account. The state equations method, the assumption of an Erlang model with fictitious traffic and the $\beta = r^{-\alpha}$ approach with constant α were examined. The two later models may be used in practice with precautions, their generalisation requires further investigations. To achieve more realistic simulation studies the behaviour of the called subscriber should be more precisely taken into account.

On the Interaction Between Subscribers and a Telephone System. 322
Myskja, A. and Aagesen, F.A.

A teletraffic system and the set of its subscribers may be considered as two distinct subsystems of a complete system including both. The two subsystems, which are of quite different natures, are interconnected across an interface, and this interconnection conveys mutual feedback signals to the two subsystems.

Previous publications have treated these phenomena from theoretical as well as practical points of view. The present paper concentrates on the fundamental statistical parameters of the subscriber subsystems and on alternative models based on these parameters. Failure rates and subscriber persistence are essential quantities, and a study of these is carried out, based on mathematical descriptions, including different effects of influence. The subscriber persistence and repetition intervals are measured for different A-subscriber categories and for different failure causes, by means of computerised data recording equipment. Different mathematical models are tested statistically versus experimental results obtained by observations on real traffic.

The Configuration Theory. The Influence of Multi-Part Tariffs on Local Telephone Traffic. 323
Kraepelien, H.Y.

Proper evaluation of usage sensitive pricing (USP) for telephone service requires knowledge of subscribers' behaviour pattern when subjected to different kinds of measured tariffs. Total local revenue generated by a subscriber population is the aggregate of payments from each individual subscriber, which - in turn - is a function of his traffic and the tariff. This paper deals with the effect of the tariff configuration on individual subscribers' local traffic. The term tariff configuration refers to the geometrical pattern formed by tariff components when the tariff is presented in a linear price-usage diagram. The first part of the paper contains a graphical presentation of how a subscriber logically reacts to different tariff configurations. In this connection the concepts of equilibrium, diminishing relative savings and incentive are introduced. Later in the paper approximate mathematical functions for the relationship between traffic and tariff are developed; i.e. demand functions for local traffic.

Some Traffic Characteristics of Subscriber Categories and the Influence from Tariff Changes. 324
Bo, K., Gaustad, O. and Kosberg, J.E.

On the basis of data recorded from the telephone network in Norway, a study upon subscribers' characteristics has been performed. Some of the characteristics have been examined for different subscriber categories, i.e. PABX lines, business and residence telephones. The influence from tariff changes upon the number of local and long distance calls, conversation time and traffic volume has been especially studied. In addition results are presented from measurements upon the subscribers' operation and reaction times.

Experimenting with the Effect of Tariff Changes on Traffic Patterns. 325
Cohen, G.

GTE Service Corporation has undertaken to experiment with exchange service pricing and to study the effect of tariff changes on teletraffic patterns as a function of a number of parameters including several demographic factors. This is part of a comprehensive study to assess the economic effect of introducing usage sensitive tariffs for local telephone service. A review of experimental design and theoretical analyses, a description of the field trial and experimental tariff, and certain general empirical results to date are presented. Observations of subscriber traffic characteristics under flat rate pricing including intergroup traffic flows as a function of time of day and the implied impact of usage sensitive pricing are also presented and discussed.

TRAFFIC ENGINEERING COMPUTATION TECHNIQUES

A General Purpose Blocking Probability Calculation Program for Multi-Stage Link Systems. 331
Kodaira, K. and Takagi, K.

An approximate internal blocking probability calculation program is described, which has been devised to reduce the procedure for practical applications and can generally be applied to an arbitrary complicated case without programming for the computation of individually derived formula. The calculation is performed by giving only a series of pseudo-instructions. The principle of the method and the functions of each pseudo-instruction are described in detail. Examples are shown to characterise this program, utilised for the traffic design of electronic switching systems development.

Probability of Blocking in Non-Hierarchical Networks. 332
Vestmar, B.

The subject of the paper is the probability of blocking in non-hierarchical networks and related design and network management aspects. The examination is based on a method of calculating the probability of blocking in a non-hierarchical network; it is assumed that (a) trunk groups between switching centres in the network are unavailable with some probability, p, and (b) the probability of a switching centre being unavailable is small and can be ignored. It is shown that the probability of blocking between any two switching centres can be expressed as a polynomial in p, where p is the probability of a trunk group (link) being unavailable. The terms in this polynomial depend on the type of network, the relative location of the two nodes, and the number of route-classes allowed (the shortest routes comprise one route-class between those two nodes, the next-shortest routes comprise another route-class, and so on). The method is applied to a grid network and practical values for the probability of blocking between nodes are obtained. These are discussed in detail and on the basis of this discussion, practical considerations relating to the design and management of non-hierarchical networks are given.

Some Applications of Quadratic Programming to the Calculation of Traffic Matrices. 333
Nivert, K. and von Schantz, C.

This paper describes three applications of quadratic programming, i.e. minimisation of a quadratic objective function subject to linear constraints, to the calculation of traffic matrices. Section 2 deals with the problem of transforming an old traffic matrix into a new one, the new row and column totals being fixed. In Section 3 we devise a procedure by which a call dispersal matrix can be transformed into a traffic matrix using partial information about the latter. Finally in Section 4 a method is developed by which a forecasted traffic matrix is obtained using forecasts on the number of subscriber per exchange and on a few large traffic streams.

The essence of the given results is that the calculated traffic matrices can be said to have the specific property of minimising a given sum of squared differences. The sum of squares is chosen so that a small value seems highly desirable from a practical point of view.

The application of these principles in the Stockholm multi-exchange area is summarised in Section 5.

An Approximate Method of Calculating Delay Distributions for Queueing Systems with Poisson Input and Arbitrary Holding Time Distributions. 334
Bear, D.

The paper describes a method of calculating approximate delay distributions in terms of the first two moments of the holding time distribution, when the standard deviation is not greater than the mean, using standard delay curves based on constant and negative exponential holding time distributions. The accuracy of the method is assessed by comparison with exact computations for the Erlangian distribution. The effect of the third moment on delay distributions is investigated in relation to the Erlangian and hyperexponential distributions.

On Point-to-Point Losses in Communications Networks. 335
Butto, M., Colombo, G. and Tonietti, A.

Some analytic models for the calculation of groups and point-to-point losses in overflow circuit switched communication networks are presented. These algorithms can be used with any routing plan and hierarchical or symmetrical networks. Stage-by-stage or conditional selection route control is adopted. Overflow and carried traffics are characterised by mean and variance. The calculation of the offered traffics takes approximately into account the groups dependence. The results obtained are compared with simulation results. The most accurate models involve a 20% relative error on losses, hence they are sufficient for most network designs and performance investigations.

Transformed Probability Distributions of Indetermined Form : Calculation Methods for Moments of Arbitrary Order and Their Application in Queueing Systems. 336
Schreiber, F.

The mathematical treatment of random processes often affords the calculation of moments M_j by means of the appropriate derivatives of a transformed distribution e.g. the Laplace transform L(s) of p.d.f. p(x). The application of this well known method might become complicated, if the transform is a fractional function $L(s) = A_0(s)/B_0(s)$ and shows at s=0 the indetermined form 0/0 which must be evaluated by the rule l'Hospital.

The indeterminateness is assumed to be of m-th order (m=0, 1, 2, ...). Based on a Maclaurin's series expansion for nominator $A_0(s)$ and denominator $B_0(s)$ a general formula and also a recursion formula are developed which allow the exact calculation of the j-th derivative.

$L^{(j)}(0)$ and thereby of the moment M_j (j=1, 2, ...). The general formula makes use of Faa di Bruno's differentiation formula. The recursion formula can be performed by a universal computer program and is used preferable for moments M_j of higher order, e.g. j 3. In order to demonstrate the recursion method in an uncomplicated case the moments M_j of the waitimg time and delay time distribution of queueing system M/G/1 are calculated for order j = 8.

Traffic Engineering with Programmable Pocket Calculators. 337
Bretschneider, G.

Programmable pocket calculators are shown to be a practical tool for the traffic engineer. The author, who has programmed the basic formulas of traffic theory for these calculators, demonstrates their usefulness by explaining some of the operating instruction charts.

TRAFFIC DISTRIBUTION IN THE NETWORK

Axiomatic Fundamentals for the Calculation of Traffic Distribution in Telephone Networks. 341
Daisenberger, G.

Let U be the total set of subscribers and A and B subsets of U. The "traffic from A to B" can then be represented by a measure y(A,B) which has the following properties : non-negative, limited, totally additive with respect to A and B.

A general formal foundation for performing calculations on traffic distributions is developed on this basis. The characteristic properties of traffic distributions as well as specific problems involved in the planning of telephone networks and practical methods of solving them are indicated in terms of this fundamental approach.

A typical example of a situation where this approach is valuable is that where telephone exchanges are to be rearranged. The traffic interrelations between the exchanges of a network, usually expressed in the form of a traffic matrix, only provide "supporting values" for the measure y(A,B). By "interpolating" between these values it is, however, possible to establish a complete measure which approximates the true measure y(A,B). In the case of a rearrangement of the exchanges, this measure provides all the necessary information on the new traffic interrelations which result.

The Effects of Non-Uniform Traffic Distribution in Switching Networks. 342
Hofstetter, H.

This paper deals with the phenomenon that in a switching network the individual link groups between any two stages do not have a uniform traffic load. Such a non-uniform load within the switching network can be due to the following causes :

External causes: When subscribers and trunks are assigned to the switching network, uniform distribution of the load over the individual matrices is not achieved.
Internal causes: The specific mode of operation of the switching control.

The effect of non-uniformly loaded link groups on the grade of service is analysed in four-stage switching networks composed of two-stage link blocks. This effect is always shown to be unfavorable. The influence of the externally caused non-uniform load on the grade of service of configurations employing the structure under consideration is intensified as the mean load of the links leaving the 1st- and last-stage matrices increases. With regard to the internal causes it is important for the switching control for outgoing traffic to be based on a suitable strategy.

Peakedness in Switching Machines : Its Effect and Estimation. 343
Heffes, H. and Holtzman, J.M.

Peakedness has been shown to have a degrading effect upon performance of switching machines. This paper reviews the effect of peakedness upon a class of electro-mechanical switching machines (describable by GI/M/N systems) and focusses on a method of estimating it.

One method of estimating peakedness is based on the measured delay; one determines which peakedness could have caused the delay. This method is desirable from the point of view of directly measuring the effect of peakedness upon performance. However, it can become unreliable when the delays are low, as they normally should be.

Another method (discussed in detail), applicable at low loads is based on within-hour samples of usage. By determining the variance-mean ratio of the number of busy servers in a GI/M/N queueing system, a peakedness estimation procedure is defined. Peakedness as a function of holding time plays a role in interpreting the results. Statistical accuracy of the procedure is discussed.

A Mathematical Model of Telephone Traffic Dispersion in Some Australian Metropolitan Networks. 344
Dunstan, A.W.

Analysis of measurements of point-to-point traffics in telephone networks in some Australian State capitals leads to a form of gravity model in which the explanatory variables are radial distance and telephone service density. The model is extended to estimate "own exchange" traffic and to include an "adjacency factor". A way is suggested of forecasting point-to-point traffic flow incorporating the influence of exchange parameters as well as measurements of initial point-to-point traffic flows.
(This paper is printed in full on pp. 84-91)

Analysis of Traffic Flows on Subscriber-Lines Dependent of Time and Subscriber-Class. 345
Evers, R.

Results of a measurement in a local exchange in Berlin (West) have been analysed. The dependence on the gained parameters from subscriber-class and time of day is discussed. As the data have been collected in one exchange only, no results concerning the influence of the local position of the observed subscribers can be presented. The structure of the traffic generated by the subscribers is described by the "call-mix" of non-blocked attempts, the grade of perseverance, the repetition-times after unsuccessful attempts, the holding times and the distribution of calls over the time of day and over the distance between calling and called subscriber. Some ideas are added in which way the results can be applied to optimisation procedures for telecommunication networks.

NETWORK AND SYSTEM RELIABILITY

Evaluation of Reliability and Serviceability in Communication Systems with its Applications to Network Planning. 411
Mori, H. and Teramura, H.

With the development of information-oriented society, communication systems have been expanding the capacity and the functional abilities, and their responsibility to the human society has become extremely large. To design and operate such systems, in addition to the conventional system engineering based mainly on efficiency maximum or cost minimum, the new engineering is required where the effectiveness of the system in the given environment is evaluated from various angles.

In this paper, the new concepts of system serviceability is discussed and the measure to evaluate it on the waiting time basis is proposed. By using this measure overall system characteristic covering service quality and reliability can be evaluated quantitatively. This measure is formulated and its relations to conventional system parameters are discussed when it is applied to the very simple system, and then the serviceability of practical telephone systems handling the hourly distributed traffic is examined. As the results, the useful empirical formula expressing the relations of proposed measure to the conventional reliability parameters are obtained. Finally several applications are shown to demonstrate the practicality of proposed measure. It is concluded that the new aspects of system engineering should be explored by introducing new concepts as proposed here in the constructive and administrative works of large-scale communication systems.

A Study of Software Reliability. 412
Andersson, H., Peram, L. and Strandberg, K.

The paper presents some new contributions to software reliability models. Terms and definitions for software reliability and associated concepts are proposed. The authors discuss the possibility of taking special properties of an SPC switching system into account. Based on this discussion they propose mathematical models, evaluate contributing factors and derive suitable characteristics.

The model validation is treated in connection with a case study.

Effects of Faults on the Grade of Service of a Telephone Exchange. 414
Kaniuk, G. and Smith, J.L.

This paper discusses the Reliability of a Telephone exchange in terms of two Stochastic processes. The first process X_t describes the fault population in the exchange in terms of the individual fault failure rates and fault durations. The second process Y_t describes the resulting changes in the grade of service induced by the fault population. The functional relationship $Y_t = g(X_t)$ is used to discuss the frequency and duration of upcrossings of Y_t in terms of the individual fault failure rates and durations on the assumption that fault durations are negative-exponentially distributed. The general theory developed is applied to an exchange with two faults.

On the Influence of Certain Typical Equipment Faults on Grade of Service. 415
Jensen, E. and Toledano, F.

This paper is concerned with the traffic effects of equipment faults causing changes in holding time distributions on a per call and device basis. In particular, the effect of so called killer devices, i.e. malfunctioning devices with relative short holding times will be studied.

The types of systems considered are (a) full availability groups with random hunting and loss, (b) link systems with loss, (c) overflow systems with faulty primary groups.

Holding time distributions are generally unrestricted however, some results are limited to neg. exp. distributions. Calls are assumed arriving according to a state dependent birth process.

Results are obtained in the form of stationary state distributions and the more important associated traffic parameters, such as carried traffics, call loss rates, trouble rates, etc. Further, a fault-checking statistical test on the number of seizures has been given.

Effectiveness Characteristics of Partly Disabled Device Groups. 416
Andersson, H. and Strandberg, K.

The paper presents formulas and methods for the determination of availability, trafficability and effectiveness characteristics of partly disabled device groups. A partly disabled device group is a group of devices, where the occupations of a subset of devices are classified as unsuccessful.

Certain fully accessible, randomly hunted loss, delay and combined loss-delay systems are treated. Formulas and methods are given for an iterative calculation of state probabilities. The methods, when programmed on a computer, give simple calculations to determine probability of call failure, call and time congestion, mean waiting times and other measures.

The methods presented are primarily intended for use in the evaluation of alternative telecommunication system design proposals.

<u>OVERFLOW TRAFFIC MODELS</u>

Some Formulae Old and New for Overflow Traffic in Telephony. 421
Pearce, C. and Potter, R.

The description of telephone traffic by channel occupancies leads, in a situation of repeated overflows, to complicated equations for the joint distribution of the numbers of occupied channels at the different stages involved. We employ a prescription through the times separating consecutive calls in the traffic stream, which enables compound systems to be considered piecemeal. Some specific formulae are derived generalising known formulae and some general questions considered through this approach.
(This paper is printed in full on pp. 92-97)

On Higher Order Moments of Overflow Traffic Behind Groups of Full Access. 422
Schehrer, R.

This paper deals with the calculation of higher order moments of traffic overflowing from groups of full access.

For the factorial moments, the ordinary moments and the central moments of such an overflow traffic, an elementary derivation is presented which does not employ a transformation by means of generating functions.

Furthermore, higher order moments of the traffic carried in finite secondary groups are considered. Exact, explicity formulae are derived for the factorial, ordinary and central moments of the traffic carried in a finite secondary group of full access.

The presented formulae for higher order moments are illustrated by numerical examples.

Mean and Variance of the Overflow Traffic of a Group of Lines Connected to One or Two Link Systems. 423
de Boer, J.

Traffic overflowing from a group of lines connected to a link network may be characterised by its mean and variance. These moments can be expressed in the Poisson traffic offered, the number of lines and the characteristic blocking quantities P_i of the network. P_i is the probability that no one of i free lines is accessible from a given free inlet of the network under given loading conditions.

The paper consists of two distinct parts. After a short survey of basic formulas, the case of two link networks is considered in the first part. Here some of the lines are connected to one network with quantities $P_{1,i}$ other lines are connected to a second network with quantities $P_{2,i}$. By a random choice it is decided whether the establishment of a connection should first be tried via the first network or via the second. For this situation the relations between the P_i of the combined network and the $P_{1,i}$ and $P_{2,i}$ are derived. It is shown that the combined network may be treated as the case of one network with quantities P_i.

In the second part of the paper approximations for the mean and variance of the overflow traffic are derived which are suitable for pocket calculators.

The Accuracy of Overflow Traffic Models. 424
Freeman, A.H.

This paper compares the equivalent random model for overflow traffic with two models based on the interrupted poisson process. One of these models is considerably more difficult to compute than the E.R. model but its accuracy is much greater and it is useful as a reference for comparing different models. The other is of comparable accuracy to the E.R. model and in some applications is more easily computed.
(This paper is printed in full on pp. 98-103)

Behaviour of Overflow Traffic and the Probabilities of Blocking in Simple Gradings. 425
Kosten, L.

This paper essentially is a condensed and modernised version of some of author's early work which is nearly inaccessible due to war circumstances. A system is studied in which several Poisson groups of sources each dispose of a number of individual servers and, moreover, jointly of a number of commons. In order to solve the problems of the loss factors in this system, subsystems are studied each comprising one group of sources with its individuals and having overflow to an infinity of commons (better called "secondaries" now, as there are no competing groups of sources). Two formulae are obtained for the distribution function of the number of occupied secondaries in those subsystems. For the original problem a (formal) exact solution is obtained as well as bounds to this solution.

Cyclic Overflow of Calls Offered to Subgroups of a Full-Availability Group of Lines. 426
Jung, M.M. and de Boer, J.

Poisson-distributed traffic is offered to a full-availability group of lines according to the following routing rule.

A group of N lines is split up into L subgroups of M=N/L lines. Likewise, the total traffic A is split up into L traffic parcels with intensity A/L. Traffic parcel no. i is offered first to subgroup no. i. Calls which are rejected by this subgroup are offered to subgroup (i+1) mod L, and those calls which are rejected again are offered to subgroup (i+2) mod L, and so on. A call is lost if it is offered to all L subgroups without being successful.

Offering rejected calls to consecutive subgroups is accomplished by the control section of the system. For determination of the load of the central control, the following quantities are of interest :

- The overflow probabilities.
- The mean number of times that a call will overflow before having success.

The present report gives methods for calculating these quantities.

Correlation Induced in Traffic Overflowing from a Common Link. 427
Wilson, K.G.

Certain links in alternative routing networks are offered independent streams of traffic from two or more sources. Previously, the behaviour of this traffic has been predicted using simulation or the equivalent random method. An analytic model for this case has been developed and can be used as a tool in investigating traffic behaviour in alternative routing networks in general.

This paper considers the moments of the traffic overflowing from the shared link and in particular the covariance. The covariance is often assumed to be zero, e.g. in simulation or else bypassed as in the equivalent random method which considers the offered traffic as if it were only one stream.

An iterative solution to the problem is given and an analytic solution to a special case, in which the offered traffic is random, is derived.
(This paper is printed in full on pp. 104-107)

SIMULATION TECHNIQUES

Statistical Problems in Simulation of Telecommunications Traffic Systems, Some Analytical Results. 431
Olsson, K.M.

A loss system in which the number of customers in the system can be described as a birth and death process with arbitrary birth and death intensities is considered. For this system we derive formulas for the variance of the number of loss calls, the proportion of lost calls and the time congestion measured by continuous observation or by different scanning methods. For the Erlang loss system an exact formula for the variance of the number of lost calls is derived for the case when the observations comprise a fixed number of calls.

For observations on the Erlang waiting system comprising a fixed number of calls, formulas are derived for the variance of the mean waiting time, the proportion of delayed calls and the mean queue length observed when calls arrive.

An Interactive Simulation System for Data Communication Network Design - ICANDO. 432
Ono, K. and Urano, Y.

This paper describes the new interactive simulation system (ICANDO) which was developed as a tool in planning and designing data communication systems and data networks. The included are study background, model and program description. This system is composed of several computer programs using FORTRAN and GPSS, associated data base and an interactive graphic display. We also demonstrate how this proposed interactive simulation system works for the problem. The paper concludes that proposed hybrid simulation models which involve both analytic and simulation techniques are highly effective for evaluating communication networks and determining the realistic design parameters with man-computer interaction.

Subcall-Type Control Simulation of SPC Switching Systems. 433
Dietrich, G. and Salade R.

The paper discusses a simulation technique for switching system control investigations which is characterised by specific simplifications in the traffic model. All call attempts are broken down into independent "subcalls" like PRESELECTION, SELECTION, ANSWER, RELEASE; dialling and signalling are represented as independent stochastic processes.

The generally applicable modelling principles for SPC type control systems, based on the subcall-type traffic model, are being described, taking the METACONTA medium size local switching system as a reference system.

Advantages, drawbacks and application range of the subcall-type simulation technique are being discussed.

Typical simulation results are presented for a 20,000 line METACONTA exchange under normal load and two overload conditions that are described by call mixes.

ENTRASIM - A Real-Time Traffic Environment Simulator for SPC Switching Systems. 435
Gruszecki, M.

The real-time environment simulator provides a model of the SPC switching system in all its complexity, using the actual processors and software. The complete switching network, all signalling and peripheral control devices and all other elements of the system are represented by memory cells in the simulator. Initially developed for software check-out in SPC systems of the Metaconta family, the environment simulator has been extended into a traffic simulation tool : ENTRASIM. This system accurately models and simulates the SPC switching machine and the external traffic environment viz. the offered calls and event sequences within each call. This new traffic simulation technique has promising applications in call handling capacity studies of SPC processors and other traffic studies where an accurate representation of the switching system and its environment is essential.

Traffic Studies of Message Switching Computers. 436
Thuan, L.D. and Bogner, R.E.

This paper studies prediction of the performance of computer controlled store and forward message switching systems, and hence the traffic levels and CPU utilisations which the systems can handle. The study is carried by both analytical and simulation methods, and the analytical results are found to be less accurate. The simulation is also used to study the dependence of response time on throughput.
(This paper is printed in full on pp. 108-115)

NETWORK MANAGEMENT AND SUPERVISION

Factors Influencing the Call Completion Ratio. 441
Riesz, G.W.

This paper begins with a brief description of network call completion statistics obtained by special processing of automatic message accounting data. In this paper "completion" is simply the ratio of answered calls to total attempts. Illustrative results are then presented for the total U.S.A. telephone system and for several individual Bell System operating companies. Seasonal variations, residual month-to-month variability, and problems resulting from abnormal traffic are discussed.

It is well known that the terminating customer's state (busy/does not answer) and the originating customer's response when a call attempt is not successful (retrial/abandonment are the major factors influencing call completion percentages. Under present U.S. service standards, these factors serve to mask network effects, including blocking and equipment failure. This paper shows that the called party class of service distribution (i.e., numbers of residence, business, multiline hunt, PBX lines, etc.) provides a basis for calculation of an expected completion rate for each central office in a large administrative area. A computer model which has been developed to identify candidate offices for potential completion improvement is described.

Procedures to detect the causes of low completion ratios and possible methods of improvement are discussed. An example of a significant shift in network completion, resulting from a tariff change, is presented, and the cost/effectiveness of completion ratio improvement activities is discussed.

Statistical Aspects in Detecting Critical Traffic Conditions. 442
Miranda, G. and Tosalli, A.

This paper aims, in the first place, at contributing to the solution of the problem of alarm criteria by attempting to demonstrate the desirability of expressing the concepts of reliability and promptness of an alarm criterion in statistical terms. This need should be borne in mind particularly when the introduction of automatic management of a network is envisaged, that is, management without the assistance of trained and expert operators.

Secondly the paper attempts to demonstrate the statistical meaning of the basic concepts of reliability and promptness by means of a very simple example, based on the observation of the traffic handled by a group controlled according to Sobel-Wald's sequential criterion (12).

Lastly, the paper tries, by the use of simulations, to show the delicacy of the assumptions which form the basis for the definition of an alarm criterion drawing attention to the possible influences, on the results deriving from the use of an oversimplified criterion, of various aspects, such as the dependence of the observations, the capacity of the group, the peakness factor, the critical level of the situations to be identified as abnormal and the transience of the abnormal situations to be identified.

Concent - An Aid to the Business Management of Telephone Networks. 443
Marlow, G. O'H.

A method to measure telephone traffic levels in remote exchanges from a central location has been developed and the technique has been applied to a medium sized metropolitan network exceeding 200,000 subscribers. Six continuous seven day (24 hour) traffic studies have been carried out on this network over a 17 month period to establish the value of, and to observe any trends in, macro network traffic parameters. This paper discusses the reasons for these studies and outlines the results obtained to this stage. An interesting outcome has been the relative stability of the weekly load factor, which has important application for estimation of call earnings.
(This paper is printed in full on pp. 116-124)

Individual Circuit Measurements. 445
Graves, R.D. and Pearson, D.A.

The measurement of usage and event data on an individual circuit basis is not new. However, recent hardware technology has made it economic to measure tens of thousands of circuits and recent mini-computer developments have greatly reduced real-time data processing costs. In 1971, this new "individual circuit/mini-computer" principle was first applied in trunking studies at New York Telephone Company; since then, many small scale trunking studies of this type have been performed throughout the United States.

This report describes the first major installation of individual circuit/mini-computer equipment for the full measurement of all COE in a large step-by-step central office (i.e. 10,000+ measurement points). The report details the advantages of software versus hardware grouping in terms of reductions in installation, administration and rearrangement costs; evaluates the benefits of individual circuit measurements (ICM) for data validation; and contrasts the capability to identify inoperative and defective plant vis-a-vis existing maintenance procedures. Finally, it summarises specific areas where individual circuit measurement techniques have improved central office performance in terms of trouble report frequencies, service observation results, and service indices.

NETWORK AND SYSTEM OPTIMISATION

Optimisation of Telephone Networks with Alternative Routing and Multiple Number of Route Group Channels. 511
Tsankov, B.

This paper deals with the optimal planning of the trunk or junction network with alternative routing, when the number of channels of each trunk group or junction group is just multiple of a number z, i.e., the number of channels is divisible by z. As a rule z is dependent on used carrier system. This is a problem of significant interest to the integrated switching and transmission (IST) network. In the paper is demonstrated the advisability of the organisation of alternative routing in such networks. An analysis of the optimising problem is presented. Calculation procedures suitable for the economical planning on a computer are achieved by means of the methods of the integer programming. Examples of the numerical results are presented.

A Comparison of System and User Optimised Telephone Networks. 512
Harris, R.J.

In a recent paper (Ref. 1) the principles of System and User optimisation were introduced and discussed for alternate routing telephone networks. (A System optimised network design is obtained by minimising the total cost of the network, subject to Origin to Destination grade of service standards. A User optimised network design is achieved by minimising the cost per erlang on chains used by each OD pair). An algorithm for determining these optimal network designs has been developed (Ref. 2) which is based upon a modification to a well known non-linear programming algorithm proposed by Wolfe (Ref. 3).

A mathematical model for dimensioning alternate routing networks developed by Berry (Ref. 6) has been used in conjunction with the optimising algorithm to obtain network designs based on the two principles. The purpose of this paper is to compare the two different network designs obtained by applying these optimising principles to the Adelaide Telephone Network.
(This paper is printed in full on pp. 125-132)

A Generalisation of Takagi's Results on Optimum Link Graphs. 513
van Bosse, J.G.

We compare the congestion in multi-stage switching networks (with random path selection) by examining the corresponding channel graphs. Takagi has shown that in a suitably chosen group one can identify an "optimum" graph, which has the lowest congestion. His proofs make use of Hölder's inequality and depend on the assumption that two or more link groups in the graphs have binomial occupancy distributions.

We present an extension of Takagi's results by verifying their validity without making any assumptions regarding occupancy distributions.

A Method for Determining Optimal Integer Numbers of Circuits in a Telephone Network. 514
Berry, L.T.M.

For a given telephone network, there exist many different junction allocations which achieve specified overall traffic congestions between each pair of exchanges. This paper considers the problem of finding a Minimum Cost network, that is, a network which satisfies the performance criterion at a minimum total junction cost. Previous models have relaxed the integrality conditions on junction numbers.
(This paper is printed in full on pp. 133-137)

On Optimal Dimensioning of a Certain Local Network. 515
Yechiali, U.

A group of m sources (telephone exchanges) is offering traffic to a group of n local exchanges. From each sources S_k (k = 1, 2, ... m) there are N_{ki} trunks leading directly to local exchange E_i (I = 1, 2, ... n). A call originating at S_k is transmitted first to local exchange E_i with probability q_{ki} (iq_{ki} = 1). If the call's destination is E_j (j = i) rather than E_i, the call is transferred from E_i to E_j.

For such a network, the optimal economic dimensioning of trunks (i.e., optimal allocation of the N_{ki}'s) is determined. It is shown that, for each source exchange, the optimal dimensioning is to direct all trunks to a single local exchange (which may differ for distinct sources). This single local exchange is determined as a function of the cost of trunks (i.e., distances) between the exchanges and the load offered by the sources to the various local exchanges.

A Probabilistic Model for Optimisation of Telephone Networks. 516
Krishnan Iyer, R.S. and Downs, T.

In this paper a probabilistic model for a telephone traffic network is developed. A system of state equations is presented for the basic building block of an alternate routing network and an explicit solution is given. From this model, the blocking probabilities at any network node may be readily obtained. As a consequence, it is possible to formulate an optimisation problem for the minimisation of the variance of the traffic arriving at the x - tandem (subject to cost and any other required constraints).
(This paper is printed in full on pp. 138-141)

DATA TRAFFIC PROBLEMS

Data Communications Through Large Packet Switching Networks. 521
Kleinrock, L. and Kamoun, F.

The topological design and adaptive routing procedure for computer networks becomes infeasible under their present form as the number of network nodes grows. In this paper we present, optimise and evaluate hierarchical procedures to be used in the case of large networks. These procedures are an extension of present schemes and rely on a hierarchical clustering of the network nodes. Models are developed to determine optimal clustering structures which lead to a minimal routing table as well as those structures which lead to a minimal computational cost for the topological design. Both optimal structures achieve enormous savings. The effect of hierarchical routing on network throughput and delay is also studied and demonstrates the efficiency of hierarchical routing for large networks.

Evaluation of Traffic Characteristics of Some Time Division Switching Networks for Data with a Plurality of Speeds. 522
Inose, H., Saito, T., Wakahara, Y. and Fukushima, Y.

Keeping uniform spacing between data samples in time-division switching network to handle data of various speed, time slot assignment mechanism can be simplified, and control memory cost can be reduced. Some time division data switching networks for large and small offices are shown and as a practically realisable solution to improve traffic handling capability, the primary rearrangement of low speed data is proposed to accommodate high speed data. The blocking probabilities for these networks without rearrangement and with primary rearrangement are derived and through numerical calculations, the effect of primary rearrangement is evaluated.

Measurement and Analysis of Data Traffic Originated by Display and Teletypewriter Terminals in a Teleprocessing System. 523
Pawlita, P.

This paper describes results of data traffic measurements at a general purpose time sharing system with moderate load of technical and commercial tasks set by the users Measurements in this field published so far deal mainly with data traffic related to typewriter or teletypewriter terminals. The present investigation treats data traffic of a collective of buffered alphanumeric display (CRT) and unbuffered teletypewriter terminals, half-duplex connected by dialling lines at a speed of 1200, 200 resp. 110 bit/s.

Using a lately available high speed hardware monitor a new method for data traffic measurement has been developed : all input and output dialogue characters of the active terminals passing the CPU-side interface of the data transmission controller were serially recorded on magnetic tape for further treatment. A special program was developed to evaluate the recorded data with regard to different time random variables dividing the dialogue cycles into characteristic periods.

Some basic differences between display and teletypewriter terminals are discussed. As one result it is shown that the mean durations of dialogue cycles of both types of terminals are approximately equal though the mean values of some other dialogue segments are quite different.

Modelling and Analysis of Store-and-Forward Data Switching Centres with Finite Buffer Memory and Acknowledgement Signalling. 524
Bux, W.

In this paper a queueing model for a data switching centre within a store-and-forward switching network is developed, in order to provide a tool for the dimensioning of such networks. The model includes the finite buffer memory and the control processor of the switching centre, as well as the data channels (transmission time plus propagation delay). Moreover, acknowledgement signalling is considered because transmitted packets are stored in buffer until an acknowledgement is received. Unacknowledged packets are retransmitted after a constant time-out.

For exponentially distributed packet lengths a closed form solution for the equilibrium distribution of queue sizes is derived. This allows to evaluate performance values of the system (buffer overflow probability, flow times etc.). For constant packet lengths a simple approach is given for estimating the most important performance values. The approximation is checked by simulation. Moreover an algorithm is proposed to analyse iteratively whole networks with the aid of the developed model.

Probability of Loss of Data Traffics with Different Bit Rates Hunting One Common PCM-Channel. 525
Katzschner, L. and Scheller, R.

This paper deals with the calculation of the probabilities of loss for data traffics with different bit rates, hunting one common data link, e.g. one PCM channel. This PCM channel is divided up into subchannels, so a superframe structure is obtained. A data channel for a data connection is constituted by allocating one or more subchannels, corresponding to the bit rate, to this connection. This allocation of subchannels to data connections is done dynamically during call establishment. According to the multiplexing techniques, two different principles of allocating subchannels to one data connection are considered : arbitrary subchannel allocation and regular subchannel allocation.

As the data traffics with different bit-rates suffer different probabilities of loss, the paper investigates means to obtain equal probabilities of loss for all types of traffic. This is done by introducing a limited availability for calls with lower bit rates. The paper concludes with a comparison between the probabilities of loss in case of the two dynamic allocation principles and with the case of permanent allocation (separate "trunk-groups for all types of traffic).

INTERNATIONAL AND I.T.U. STUDIES

Traffic Engineering in Developing Countries. Some Observations from the ESCAP Region. 531
Elldin, A.

Conditions for traffic engineering in developing countries as regards difficulties in satisfying an ever increasing demand for telecommunication equipment services. Available technical and human resources to face the problem.

The problem quantity or quality reflected in some typical overloading problems. The importance of network planning and study of subscriber dialling behaviour.

Formulation of how traffic engineers from the developed world best can help the developing countries.

International Network Planning - Two Case Studies. The Pan-African Telecom. Network and the Middle East and Mediterranean Telecom. Network. 532
Engvall, L.

During the period 1968-1976 two projects on international network planning have been undertaken through the International Telecommunication Union. The Pan-African Telecommunication Network, embracing 30 countries in Africa (principally south of Sahara), was the first to make extensive use of computer calculation methods. These were based on the principles of alternative routing with economic comparison, explained in the CCITT GAS 2 Manual. These principles refer to a local network of multi-exchange configuration and had to be adapted to the case of an international network for a large region.

The other project covers a region including the 20 Arab countries together with seven other neighbouring countries in Europe and Africa. In this study the method previously developed for the Pan-African network has been refined to include a large variety of cost functions for possible transmission media to be selected on the most economical annual cost per circuit basis. The effect of time differences on international calls has been considered by applying to the traffic input the distribution function over 24 hours as presented in ITC papers in the congresses in Munich and Stockholm. In the selection of the most economical international network, the introduction of a dedicated satellite system for the project region has been considered as well.

Integration of Communication Satellite into the International Networks; Planning and Economy, Traffic Engineering. 533
Pernau, W. and Urmoneit, W.

Today, there is a world-wide competition between cable and satellite systems as reliable transmission media. Ten years of experience in the operation of communication satellites allow more accurate statements to be made on an optimum use from the viewpoint of planning, operation and economic efficiency.

Sections 1 and 2 of this paper include not only a brief report on the growth rate of the telephone stations and the development of the intercontinental traffic but also a description of the operational and economical application of cable and satellite systems to maintain an efficient intercontinental telecommunication service.

Section 3 deals with the maritime satellite (MARSAT) system which is at present under development. It is shown that in the case of multiple access to the N satellite channels (all shore stations have access to the N satellite channels) a proper technical solution (no dual seizures on the operating and signalling channels) is only possible when the switching system is designed as a delay and loss system. Fig. 3.5-1 shows the allocation of the number of waiting positions to the number of satellite channels with given traffic parameters.

The aim of the study outlined in Section 4 is a comparison between the total costs of a satellite system and the costs of a terrestrial network which would meet the same demand for circuits but with conventional means. The annual costs per circuit kilometer, which are a specific value for each network, offer themselves for the cost comparison.

Structuring of a Telephone Network with Alternative Routing but with no Hierarchisation of the Centres. 534
Chapuis, R.J., Hofer, C. and Perrin, J.L.

This study is part of a more general one intended to compare the costs of an international network with a given grade of service between any pair of international centres in two distinct situations:

(i) where the network is designed with systematic use of alternate routing and without regard to any particularism stemming from national considerations;

(ii) where a network is based on the sum of the decisions peculiar to each of the countries it connects, the decisions being guided by the desire on the part of each country to maximise the financial return derived from its international service.

These two policies for planning the international network are likely to lead to very divergent results in terms of the structure of the network. A comparison between the North American network and the international network in Europe is instructive in this respect.

This is a theoretical study, but a sufficiently representative model of an international network has been used. The model of the international network is represented by a square matrix with five rows and five columns (i.e. 25 countries). A non-hierarchical network has been assumed.

A revision of CCITT Recommendation Q.13 is to be adopted for study by the VIth Plenary Assembly and the present theoretical study could well contribute in this context towards clarifying certain points concerning the structuring of a non-hierarchical network.

LINK SYSTEMS

Wide-Sense Non-Blocking Networks, and Some Packing Algorithms. 542
Smith, D.G. and Rahman Khan, M.M.

It is demonstrated that for a three-stage symmetrical network, with one link between switches in successive stages (i.e. (m, n, r)), the number of middle switches required for non-blocking cannot be less than $2n - \frac{n}{r}$.

In addition, results showing the effect of a number of packing rules on the internal congestion of a particular network are presented.

Blocking Probabilities for Connecting Networks Allowing Rearrangements. 543
Hwang, F.K.

We consider connecting networks in which we allow existing calls to be rearranged (rerouted) to accommodate a new call. This capability exists, for example, in the network of the recently proposed automatic main distribution frame [(2, 3, 4)] (where the need for switching live calls does not arise). For practical reasons, we are particularly interested in networks in which we allow a fixed small number, say t, of calls to be rearranged. The probability that a (random) new call is still blocked even allowing t-1 rearrangements is defined as in the t^{th}-trial-blocking probability. When t is unlimited, then the network is said to allow total rearranging. If a network can accommodate any new call under total rearranging, it is said to be rearrangeable.

In this paper, we propose a model for computing the t^{th} - trial-blocking probability recursively for the multi-stage Close network. The computational complexity of this model is discussed and approximations suggested. We also give various modifications of this model to fit particular networks and possible extensions to more general networks. Finally, some simulation results and computer implementations of this model are mentioned.

Some Blocking Formulae for Three-Stage Switching Arrays with Multiple Connection Attempts. 545
Morrison, M. and Fleming, P. (Jr.)

The performance of three stage link matrices is improved by multiple marking control techniques by which the various third stages are searched for an idle outlet of the proper class. Matrices with low occupancies on the inlets show greater improvements, but the most impressive gains are found when the primary stages are square, or provide expansion into the second stage.

Design of Mixed Analogue and Digital Switching Networks. 546
Harland, G.

Modern common control switching systems invariably use link trunking within individual exchanges whereas step-by-step operation is still the normal way of working in complete multi-exchange networks. The introduction of common-channel signalling by means of inter-processor data links, together with integrated digital switching and transmission, provides a convenient opportunity to review the arrangements for trunking multi-exchange switching networks.

An approach to the design of efficient link trunking arrangements and its application to multi-exchange switching networks is described. The impact of introducing digital switching into such networks is considered and possibilities for providing efficient multi-exchange networks having blocking probabilities substantially independent of the number of switching stages, are discussed.

Congestion in a Switching Network with Rearrangement. 549
Boyd, J.C. and Hunter, J.M.

This paper extends the work of Akiyama and Ershov to determine the probability of blocking for several re-arrangement algorithms. Point-to-route congestion, but point-to-point re-arrangement is considered in a three stage Close network. The distribution over the route is obtained by the use of a loss function. Simulation and analytical results are given for each case over a range of traffic.

Point-to-Point Selection versus Point-to-Group Selection in Link Systems. 541
Lotze, A., Roder, A. and Thierer, G.

This is the third paper of a three paper study presented at the 8th ITC.

The first paper /1/ derives the new and reliable PPL-method for the calculation of the point-to-point loss in link systems and presents guidelines for the dimensioning of crosspoint saving link systems operating in the point-to-point selection mode (PPS-mode). The second paper /2/ shows that folded and reversed systems (so-called one-sided systems) can be mapped into load and loss equivalent two-sided ones. Thus, the calculation of loss for point-to-point selection (PPS) mode as well as for the point-to-group selection (PGS) mode can be performed by means of the method PPL /1/ or the method CLIGS /3,4/, respectively. The aim of this third paper is to give the designers and field engineers criteria how to judge and to compare some important properties of link systems operating either in the PPS-mode or PGS-mode.

Investigations on Folded and Reversed Link Systems. 544
Lotze, A., Roder, A. and Thierer, G.

This paper is the second of a three paper study presented at the 8th ITC. The first paper deals with the point-to-point loss in two-sided link systems /1/ . The third one gives a comparison between the point-to-point selection mode vs. the point-to-group selection mode /2/ .

This paper here deals with one-sided link systems. These systems can be mapped into equivalent two-sided ones with respect to their traffic behaviour. Thus, methods for the approximate calculation of loss in case of the point-to-point selection mode as well as for point-to-group selection mode can be applied. A good agreement between simulation and calculation is achieved in both cases.

"PPL" - A Reliable Method for the Calculation of Point-to-Point Loss in Link Systems. 547
Lotze, A., Roder, A. and Thierer, G.

This paper belongs to a three paper study presented at the 8th ITC; the other two papers deal with one-sided link systems /1/ , as well as with the comparison between the point-to-point selection mode versus the point-to-group selection mode /2/ .

This paper here, being the basis for the other two papers, presents the new PPL-method for the calculation of the point-to-point loss in two-sided link systems. It uses quite a new way of solution basing on the derivation of an effective accessibility from a starting to a destination multiple. The calculated results for many various structures of link systems with S=3, 4, 5 and 6 stages are in good agreement with results obtained by simulation. Regardless of the easy programmability, a selection of design diagrams is included. They allow the direct design of a crosspoint-saving link system with prescribed number of lines, carried traffic, and PP-loss by reading off one set of parameters from the diagram.

S.P.C. SWITCHING SYSTEMS I

Traffic Calculations in SPC Systems. 611
Villar, J.E.

One of the problems that nowadays demands considerable attention from traffic engineers is the determination of the call handling capacity of SPC switching systems. This paper discusses four examples of new traffic problems that have arisen with this new generation of switching systems, and then addresses the problem of defining processor capacity and discusses several approaches to the dimensioning of a multiprocessing system.

A survey is given of the main analytical methods currently employed for estimating the capacity of an SPC system, in which both non-probabilistic and probabilistic methods are covered. Following this, a brief discussion is given of the principal techniques of system simulation currently used to estimate SPC capacity.

Investigations on the Traffic Behaviour of the Common Control in SPC Switching Systems. 612
Weisschuh, H. and Wizgall, M.

This paper deals with the investigation of the traffic behaviour of the common control of an experimental PCM switching system with stored program control.

An outstanding characteristic feature of the switching system is extensive preprocessing of control information by peripheral control units.

The investigations are done by simulation and calculation, resp. For simulation a model of the entire switching system is developed which includes a submodel for the subscriber behaviour.

For calculation a simplified model of the common control is developed and analysed. This model consists of a single server system with batch input (variable batch size) at equi-distant instants and service times according to an Erlangian probability distribution function. Each arrival causes a constant overhead phase in the server.

Numerical results (obtained by simulation) show the influence of the subscriber behaviour and of the preprocessing functions on the performance of the system. Furthermore, interesting traffic values for the switching system are given.

The Waiting-Time Distribution for Markers and Other Peripheral Devices in SPC Switching Systems. 613
Gerrand, P. and Guerriero, A.

The waiting-time distributions for calls at markers (and other peripheral devices) are obtained for three design strategies : direct interrupt, periodic polling and complex polling of the marker. Analytical results are already applicable to the direct interrupt case, using the Crommelin model. The periodic polling case requires fresh analysis fully described in this paper, to take into account the limited accessibility of the program serving the marker queue. The case of complex polling was found to be too difficult for direct analysis, and a simulation program was implemented to obtain the desired results. The simulation model also serves as a check on the analytical results for the periodic polling and direct interrupt cases.

Conclusions are given concerning the values of the critical parameters that lead to significantly different waiting-time distributions in the design alternatives.

S.P.C. SWITCHING SYSTEMS II

Methods of Estimating Central Processing System Traffic Performances in SPC-Electronic Switching Systems. 621
Itoh, M., Nunotani, Y., Ueda, T. and Okada, K.

This paper describes suitable traffic performance estimation methods for each switching program design phase : fundamental traffic investigation in basic study of switching program design, newly developed hybrid simulation technique in the fundamental design phase, full-scale simulation technique in the detailed design phase and office load test in an actual system. These have been developed and gradually made better in accordance with the advance in designs of electronic switching systems (ESS) with stored program control (SPC) in NTT. This paper explains in particular detail hybrid simulation technique and full-scale simulation technique with respect to a way of composing a simulator, manpower and time necessary for its composition, program size, accuracy and flexibility.

Traffic Problems within Mulitprocessor SPC Systems. 622
Bradley, I.A. and McTiffin, M.J.

The particular traffic problem discussed is that of store contention. This is a characteristic displayed by multiprocessor systems when several processors compete for access to the store modules. It has the effect of reducing the processing power of large systems below that which might be expected by extrapolation from smaller systems.

An approximate method has been developed for calculating the effect of store contention. The queueing time at each store module modifies the arrival rate of store accesses which in turn affect the queueing times. Thus the solution is found from a set of simultaneous non-linear equations which may be solved by iteration.

Priority Models for Communication Processors Including System Overhead. 623
Herzog, U.

Various scheduling strategies for communication processors are discussed and uniformly described by means of the so-called "Pre-emption Distance".

Then, two types of communication processors are described: processors with fully automatic (hardware) priority-interrupt systems and, secondly, processors with software controlled priority-interrupt systems. For both, queueing models including Pre-emption Distance Priorities are presented and analysed.

Numerical results show how software overhead influences the characteristic performance values.

The paper contains the following sections :

1. Introduction.
2. Classification of Pre-emption Distance Priority Strategies.
3. Modelling of Communication Processors.
4. Analysis.
5. Numerical Results.
6. Summary and Outlook.

Analytical Study of Feedback Effects on Processor Traffic in SPC Systems. 624
Haugen, R.B., Jensen, E. and Sanchez-Puga, M.J.

This paper is concerned with the interdependence between various tasks which have to be performed by the processors in an SPC system.

Principles of simplifications of the queueing processes, which will allow investigations by analytic tools, are presented for a one-processor system.

Further, the analytical technique has been demonstrated on a special problem of estimating the waiting time distribution for tasks with low interrupt priorities.

STATISTICAL ANALYSIS OF TRAFFIC DATA

Studies on the Probability of a Called Subscriber Being Busy. 631
Lind, G.

The paper treats some models intended to describe the losses of the traffic towards a group of subscribers, where the losses are caused by the called subscriber being busy.

First a model is developed describing the joint distribution of the total number of incoming calls and the number of lost incoming calls for a single subscriber. The model is Markovian. The means, the variances and the covariance of this distribution are derived.

Then a method to perform calculations for a whole group of subscribers is given, based on the model for a single subscriber. The random quantity considered is the proportion of incoming calls to the group of subscribers meeting with busy subscriber. A method is given to obtain the mean and the variance of this proportion. Some special cases are dealt with.

The loss in question is usually much greater than the mean traffic per subscriber.

In measurements we have found, as an example, a mean traffic per subscriber of about 0.04 erl., while the proportion of calls meeting with busy subscribers amounts to 15 - 20%. The main explanation of this seems to be the skewness of the distribution of the subscribers' traffic.

Finally is given a brief account of a theory of the same principal kind as above, except that it is based on a non-Markovian model for the single subscriber.

A Non-Constant Failure Rate Distribution in the Conversation Time of Telephone Calls. 632
Nombela, L.F. and Pillado, J.M.

Traditionally, in traffic theory, it is assumed that the negative exponential law is a good mathematical model to estimate and approach the conversation time of both, local and trunk telephone calls.

Its relative simplicity has been the main reason for its acceptance in the conventional dimensioning methods.

Recently, the introduction of multiple and sophisticated metering equipment in exchanges has provided information enough to study the statistical law of this important traffic parameter.

From several samples, with more than 20,000 local calls each, taken from cross-bar exchanges, and more than twenty samples, with more than 1,000 calls each, taken from different kind of trunks in the Spanish Telephone Network it has been tested with satisfactory results, that the Weibull distribution is a much better model than the exponential negative one to determine this important traffic parameter. In this paper, different methods to estimate the three parameters defining the Weibull distribution are discussed. Here is also included a computer programme writing in FORTRAN IV to calculate these parameters.

The Variance of Observations on Markov Chains. 633
Songhurst, D.J.

This paper concerns the variance of observations on discrete parameter Markov chains. It is assumed that observations are made at a number of consecutive epochs, each observation being a function of the state of the system at that epoch, and it is required to find the variance of the sample mean of the observations. Under a geometric ergodicity condition on the Markov chain the asymptotic form of this variance is shown to satisfy a relatively simple equation which is solved for a range of practical cases. The theory can most usefully be applied to systems carrying telephone traffic by assuming observations at the instants of call arrivals, and therefore giving results applicable to congestion measurements. The main limitation of this application of the theory is the need to assume exponentially distributed call holding times.

Statistical Design of Load Balance Traffic Studies. 634
Pedersen, O.

This paper deals with the experimental design problem of deciding how much data is needed for a traffic study to determine whether or not dial office administration objectives related to the maintenance of load balance of equipment groups are being met with an assurance that the risks of drawing wrong conclusions from the data are under control. Procedures are presented for planning tests based on the use of the studentized range of the hypothesis that the range of equipment group loads or other traffic characteristics lies within prescribed limits. Least squares estimates of the standard deviation of random effects are obtained by use of analysis of variance techniques. Range methods of estimating the standard deviation are also presented for the purpose of reducing the data processing needed for a study. Examples are given to illustrate the methods.

Call Generation, Holding Times and Traffic Load for Individual Telephone and Telex Subscriber Lines. 635
Anderberg, M. and Wikell, G.

With a minicomputer equipment, jointly owned by the Swedish Telecommunications Administration and the L.M. Ericsson Telephone company, extensive measurements have been carried out on individual subscriber lines.

This paper aims at describing the fundamental parameters for traffic, calls and holding times and their interaction. Some traffic profiles are given and the paper gives a description of how unsuccessful call attempts are repeated in telex traffic. Attention is also drawn to the fact that during a busy hour only about 20-30% of residential telephone subscribers are active.

A Modified Chi-Squared Test for Computer Simulations of Telephone Switching Networks. 636
Peterson, M.M.

A problem that arises in assessing the results of computer simulations of telephone switching networks and in comparing them with simplified mathematical models of the same networks is the determination of the equilibrium distribution of the number of busy circuits in a group which may be a small part of a complex whole.

The traditional means of determining whether or not a particular distribution, such as the Erlang, the Engset or the Bernoulli may be considered appropriate has been to take single observations of a simulation so widely spaced in time as to be at least approximately independent. These observations are then used to perform a traditional chi-squared test of goodness of fit to the proposed distribution.

In the paper it is shown that if the total times during a simulation run in which r out of a group of n devices are busy are observed (for r = 0, 1, 2, ..., n) and if standard assumptions are made about the nature of the stochastic process of call arrivals and terminations, then a simple algebraic function of these observations has an approximate chi-squared distribution and may be used to test hypotheses about such distributions in the same way as the traditional test. The new test makes use of the whole time of the simulation run and can lead to a marked saving in the run length needed in order to reach a decision.

REPORTS

Report of the Working Party for the Documentation of the First Six International Teletraffic Congresses.
de Lange, S.J. 641

This report briefly reviews the activities of the Documentation Working Party between January 1973 and May 1976. The main task was to select 10 reference libraries in the world which would be prepared to hold complete sets of papers from all ITC's and supply copies to bona fide research workers. This has been done and the co-operating libraries have been supplied with complete sets of past ITC proceedings. The names and addresses of these libraries are given in Appendix B of the report; also included (Appendix A) is the financial statement covering the whole six years of the Working Party's operations.

Report of the Working Party on Teletraffic Training.
Elldin, A. 642

This report reviews the activities of the Training Working Party since the last ITC and also gives a brief survey of traffic engineering courses and seminars in which ITC delegates participated as lecturers. The report notes the demand for teletraffic training in the developing countries and suggests how to meet it.

Equivalent High Usage Circuits and Their Application in the Analysis and Design of Local Networks Employing Alternate Routing

J. S. HARRINGTON

Telecom Australia, Perth, Australia

ABSTRACT

The cost of routing traffic between origin and destination exchanges employing alternate routing can be expressed in terms of the direct route alone by adding to it a number of circuits to equal the cost of routing the traffic, overflowing from the direct route, over the alternate routes of the network. EQUIVALENT HIGH USAGE CIRCUITS, the sum of the actual and additional direct route circuits, are shown to be functions of offered traffic, marginal occupancy (sometimes referred to as cost factor) and efficiency of the traffic switching machine. Tables of equivalent high usage circuits, for a range of marginal occupancy values and covering both full and limited availability trunking with various link losses, are presented. Examples are given in the use of the tables for optimisation of availability, comparisons of switching equipment and local network analysis. Curve fitting equations, relating actual and equivalent high usage circuits to pure chance offered traffic with marginal occupancy as parameter, are given for full availability trunking.

1. INTRODUCTION

The alternate routing of telephone traffic in local networks is a practice widely employed by telephone administrations in providing communication between telephone exchanges in the most economic manner. As the name implies more than one traffic route is generally provided between origin and destination exchanges. Calls are offered first to the most economic route and when it is congested they overflow onto an alternate route; further alternate routes selected in pre-determined sequence may be provided.

Alternate routing design is concerned with determining the number of circuits on each traffic route to achieve a minimum cost solution in switching each origin/destination traffic parcel. To this end alternate routing optimising equations, that allow the network to be designed in an orderly manner, have been established.

These equations specify a traffic property of the direct route that enable the number of circuits on the route to be determined. As the number of direct route circuits will result in the minimum cost solution of switching the origin/destination traffic parcel, it can be reasoned that the routing cost should be known at this point rather than at the completion of dimensioning the network when alternate route costs are apportioned to the direct routes. By combining the direct and alternate route circuits to form an equivalent number of direct route circuits (EQUIVALENT HIGH USAGE CIRCUITS) the routing costs can be expressed as a single cost.

There are considerable advantages in having the routing cost as a single component because, when comparing the economic advantages of one type of switching equipment against another or of different network designs (Tandem exchanges — number and location), it is the total cost of switching each traffic parcel that is important. Without exception it will be found in these cost comparisons that the cost variation in the alternate route cost component is significantly larger than the cost variation in the direct route cost component.

The parameters used in the calculation work of this paper were chosen to suit the link trunked Ericsson ARF 1GV Equipment, both two and three stage. The results can be applied however to other link trunked switching equipment and to full availability switching machines.

2. ALTERNATE ROUTING OPTIMISING EQUATIONS

In the derivation of alternate routing optimising equations the method presented by Dr C.W. Pratt (REF. 1) is followed closely.

Consider the alternate routing pattern of Fig. 1. The diagram shows only that segment (of a whole network) which takes part in the routing of traffic from an origin I to a destination J. Route 1 is the direct route and as it is the most economic route the IJ traffic parcel is offered first to it. Route 2 is the first alternate route and has offered to it the overflow traffic from route 1. Routes 3, 4 and 5 are final routes and are dimensioned for low congestion.

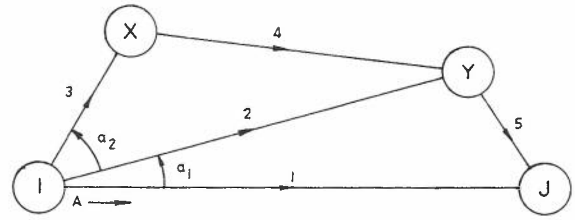


FIG. 1 ALTERNATE ROUTING PATTERN

Assume A is the pure chance IJ traffic parcel offered to N_1 circuits on route 1 and overflowing traffic a_1, which, along with other overflow traffics from direct routes from origin I to other destination exchanges parented on tandem exchange Y, is offered to route 2. Not all of this traffic is carried by route 2 and a_2 is that component of a_1 which overflows and is offered to route 3. As routes 3, 4 and 5 are dimensioned to provide good grades of service traffic a_2 can be assumed to be carried by routes 3 and 4 and traffic a_1 by route 5. Circuit quantities N_2, N_3, N_4 and N_5 on routes 2 to 5 respectively are required to carry the IJ overflow traffic component. The circuit quantities N_1, N_2............N_5 are considered to be continuous variables and may therefore be non integer. The cost of routing the IJ traffic parcel is given by,

$$C = C_1N_1 + C_2N_2 + C_3N_3 + C_4N_4 + C_5N_5 \quad \text{(1)}$$

where C is the total cost and C_1, C_2...........C_5 are the costs per circuit on routes 1, 2............5 respectively.

Assuming a minimum cost solution exists any change in N_1 or N_2, effecting changes in N_3, N_4 and N_5 (congestion standards to be maintained), increases the routing costs.

Consider that part of Fig. 1 bounding routes 2, 3 and 4 as shown in Fig. 2. The overflow traffic a_1 is offered to N_2 circuits and traffic a_2 overflows onto routes 3 and 4.

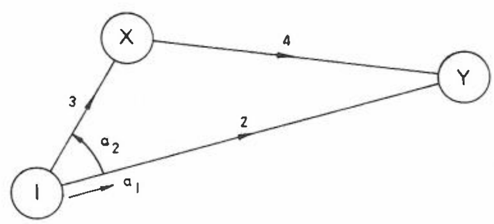


FIG. 2 PART OF ALTERNATE ROUTING PATTERN OF FIG. 1

The cost of routing the offered traffic a_1 from I to Y is given by,

$$\text{COST} = C_2N_2 + C_3N_3 + C_4N_4 \quad \text{(2)}$$

The cost is a function of one independent variable N_2 since if N_2 is chosen, N_3 and N_4 are adjusted to maintain the congestion standards.

For the cost to be a minimum the rate of change of cost with respect to N_2, a_1 being fixed and N_2 being $>$ 0, must be zero.

$$\text{Hence} \frac{\partial COST}{\partial N_2} = 0 = C_2.1 + C_3\left(\frac{\partial N_3}{\partial N_2}\right) + C_4\left(\frac{\partial N_4}{\partial N_2}\right) \quad \text{.......... (3)}$$

$$\therefore \quad 0 = C_2 + C_3\left(\frac{\partial N_3}{\partial a_2}\right)_{E_3}\left(\frac{\partial a_2}{\partial N_2}\right)_{a_1} + C_4\left(\frac{\partial N_4}{\partial a_2}\right)_{E_4}\left(\frac{\partial a_2}{\partial N_2}\right)_{a_1} \quad \text{.......... (4)}$$

Where E_3 and E_4 are the grades of service of routes 3 and 4 respectively.

The following quantities are now defined for a route of N circuits offered traffic A and carrying traffic A_c at a grade of service E :

Marginal Occupancy = $H = \left(\frac{\partial A_c}{\partial N}\right)_A = -\left(\frac{\partial a}{\partial N}\right)_A$

where a is the overflow traffic ($a = A - A_c$)

The marginal occupancy of a route is the rate of change of the traffic carried with respect to the number of circuits when the offered traffic is held constant.

Marginal Capacity = $B = \left(\frac{\partial A}{\partial N}\right)_E$

The marginal capacity of a route is the rate of change of the offered traffic with respect to the number of circuits when the grade of service is held constant.

Equation (4) then becomes,

$$0 = C_2 + C_3\frac{1}{B_3}(-H_2) + C_4\frac{1}{B_4}(-H_2)$$

$$\therefore \frac{C_2}{H_2} = \frac{C_3}{B_3} + \frac{C_4}{B_4} \quad \text{.......... (5)}$$

This is an optimising equation as, in H_2, it determines the rate of change of carried traffic with respect to route 2 circuits. The marginal capacity B is dependent more on the grade of service standard and the efficiency of the switching machine than the amount of offered traffic (see Section 5) and so over a small range of offered traffic B can be considered constant. This leads to two important results:

(i) The value of H_2 can be estimated fairly precisely.

(ii) $N_3 = a_2/B_3$ and $N_4 = a_2/B_4$ (The accuracy is quite good because a_2 is small.)

Equation (2) then becomes,

$$COST = C_2N_2 + C_3\frac{a_2}{B_3} + C_4\frac{a_2}{B_4}$$

$$= C_2N_2 + a_2\left(\frac{C_3}{B_3} + \frac{C_4}{B_4}\right)$$

$$= C_2N_2 + a_2\frac{C_2}{H_2} \text{ (from equation 5)}$$

$$= C_2\left(N_2 + \frac{a_2}{H_2}\right) \quad \text{.......... (6)}$$

The dimensions of a_2/H_2 = circuits and the term

$\left(N_2 + \frac{a_2}{H_2}\right)$ becomes equivalent route 2 circuits = N_{2eq} which is a continuous variable.

As the term direct route does not distinguish between a route dimensioned to a grade of service standard or to a cost factor criterion, routes dimensioned to the latter will be termed high usage routes. Routes 1 and 2 of Fig. 1 and route 2 of Fig. 2 are therefore high usage routes. N_{eq} is then titled 'Equivalent High Usage Circuits'.

Using N_{2eq} in place of Fig. 2 reduces Fig. 1 to,

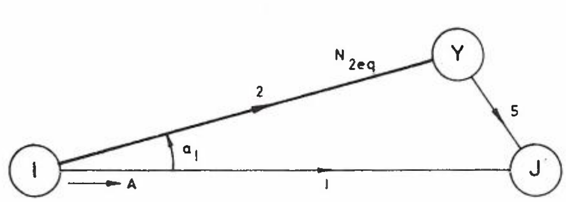


FIG. 3 SIMPLIFIED ALTERNATE ROUTING PATTERN OF FIG.1

Cost equation (1) reduces to,

$$C = C_1N_1 + C_2N_{2eq} + C_5N_5 \quad \text{.......... (7)}$$

The cost C is a function of one independent variable N_1 since if N_1 is chosen N_{2eq} and N_5 must be adjusted for the change to the overflow traffic a_1.

For C to be a minimum and N_1 being > 0,

$$\frac{\partial C}{\partial N_1} = 0 = C_1.1 + C_2\left(\frac{\partial N_{2eq}}{\partial N_1}\right) + C_5\left(\frac{\partial N_5}{\partial N_1}\right)$$

$$\therefore \quad 0 = C_1 + C_2\left(\frac{\partial N_{2eq}}{\partial a_1}\right)_{H_2}\left(\frac{\partial a_1}{\partial N_1}\right)_A + C_5\left(\frac{\partial N_5}{\partial a_1}\right)_{E_5}\left(\frac{\partial a_1}{\partial N_1}\right)_A \quad \text{.......... (8)}$$

A new quantity appears in this expression and will now be defined.

Equivalent Marginal Capacity = $B_{2eq} = \left(\frac{\partial a_1}{\partial N_{2eq}}\right)_{H_2}$

The equivalent marginal capacity of a high usage route is the rate of change of the offered traffic with respect to the equivalent number of high usage circuits when the marginal occupancy is held constant.

Equation (8) then becomes

$$0 = C_1 + C_2\frac{1}{B_{2eq}}(-H_1) + C_5\frac{1}{B_5}(-H_1)$$

$$\therefore \frac{C_1}{H_1} = \frac{C_2}{B_{2eq}} + \frac{C_5}{B_5} \quad \text{.......... (9)}$$

This also is an optimising equation as it enables H_1 to be determined. B_{2eq}, the equivalent marginal capacity of route 2, is more dependent on the value of H_2 and the efficiency of the switching machine than the amount of offered traffic (see Section 3) and can therefore be estimated fairly accurately. H_1 can be estimated to a good degree of accuracy and,

$N_{2eq} = a_1/B_{2eq}$ and, as before $N_5 = a_1/B_5$

Equation (7) then becomes

$$C = C_1N_1 + C_2\frac{a_1}{B_{2eq}} + C_5\frac{a_1}{B_5}$$

$$= C_1N_1 + a_1\left(\frac{C_2}{B_{2eq}} + \frac{C_5}{B_5}\right)$$

$$= C_1N_1 + C_1\frac{a_1}{H_1} \text{ (from equation 9)}$$

$$= C_1\left(N_1 + \frac{a_1}{H_1}\right) \quad \text{.......... (10)}$$

$$= C_1N_{1eq} \quad \text{.......... (11)}$$

Using N_{1eq} Fig. 3 is reduced therefore to a single high usage route of N_{1eq} circuits. N_{1eq} is a function of A, H_1 and the efficiency of the switching machine.

The result of equation (11) is a very important one for it enables the cost of routing the IJ traffic parcel to be determined from A, H_1 and the efficiency of the switching machine.

The optimising equations for the alternate routing pattern of Fig. 1 are given by equations (5) and (9). These are,

$$\frac{C_2}{H_2} = \frac{C_3}{B_3} + \frac{C_4}{B_4} \quad \text{and} \quad \frac{C_1}{H_1} = \frac{C_2}{B_{2eq}} + \frac{C_5}{B_5}$$

Two alternate routing patterns of greater complexity as used in the Australian local network are briefly studied in Appendix 1.

3. CALCULATION OF EQUIVALENT HIGH USAGE CIRCUITS

The Geometric Group method (Ref. 2) was used to calculate the moments of the overflow traffic for both the full availability case and a range of limited availabilities with various values of link congestion. The availabilities chosen were 80, 60, 40, 30, 20, 10 and 5; for each availability seven loss probability values were chosen to suit the Ericsson Group Selector (ARF—GV).

Inlet loadings of 0.5, 0.6, 0.7 and 0.8 erlang/inlet for the two stage GV and link ratings of 50, 60 and 70 for the three stage GV result in probabilities of meeting internal blocking in accessing the last free outlet in a group, not employing interconnecting or commoning, of,

$\frac{1.0}{3}$, $\frac{1.2}{3}$, $\frac{1.4}{3}$ and $\frac{1.6}{3}$ respectively for the two stage GV, and

$(0.5)^5$, $(0.6)^5$ and $(0.7)^5$ respectively for the three stage GV

With the three stage GV, as the probability of internal blocking is a function of the average link occupancies, b and c, of the second and third stages respectively (internal blocking = f (b + c - bc)), it is convenient to use a single term 'Link Rating' to describe the blocking probability. (Link Rating = 100 (b + c - bc)).

Although of special significance to the ARF–GV the values of loss probability and the associated calculations are of general use and can be applied to other link trunked switching machines. Switching machines employing grading but without link loss are also covered.

Within each availability range nine circuit values were chosen. These were related to the availability, being,

0.6, 0.8, 1.0, 1.4, 1.8, 2.2, 2.8, 3.4 and 4.0 times respectively the value of availability.

For each combination of availability, loss probability and circuit value offered traffics necessary to give marginal occupancy values of 0.1 to 0.8 in steps of 0.1 and grades of service of 1/200 and 1/500 were determined. The method of calculation, by computer, was to range the offered traffic to first encompass the values of marginal occupancy and grade of service and then determine the required value of traffic by a method of successive approximations. In Table 1, which shows a sample of the computer output, MQ is the availability and P the Geometric Group parameter.

TABLE 1 DETERMINATION OF A FOR SET VALUES OF H AND G.O.S.

MQ = 20	N = 20	LINK RATING = 70				
FOR N=19,20 & 21	P=0.1530,0.1681 & 0.2077				REQ'D VALUES	
A	a,N=20	∂a,N=21	∂a,N=19	H	H	A
34.00	15.2278	-0.8469	0.8804	0.8637	0.10	12.69
31.00	12.4353	-0.8157	0.8541	0.8349	0.20	14.40
28.00	9.7189	-0.7714	0.8163	0.7939	0.30	15.87
25.00	7.1223	-0.7063	0.7600	0.7332	0.40	17.34
22.00	4.7223	-0.6083	0.6733	0.6408	0.50	18.98
19.00	2.6513	-0.4629	0.5391	0.5010	0.60	21.00
18.60	2.4101	-0.4395	0.5167	0.4781	0.70	23.77
16.80	1.4556	-0.3250	0.4033	0.3642	0.80	28.38
15.20	0.8126	-0.2182	0.2897	0.2539	GOS	
13.80	0.4223	-0.1330	0.1908	0.1619	1/200	10.84
12.40	0.1830	-0.0670	0.1060	0.0865	1/500	9.81
11.20	0.0740	-0.0305	0.0533	0.0419		
10.20	0.0294	-0.0133	0.0255	0.0194		
9.20	0.0097	-0.0048	0.0101	0.0074		

The marginal occupancy value - the rate of change of carried traffic with respect to route circuits, the offered traffic being constant - was determined by averaging the change in overflow traffic one circuit either side of the nominal value. The curve of Fig. 4 was found to be typical and the marginal occupancy calculated by this method is in close agreement with the correct value.

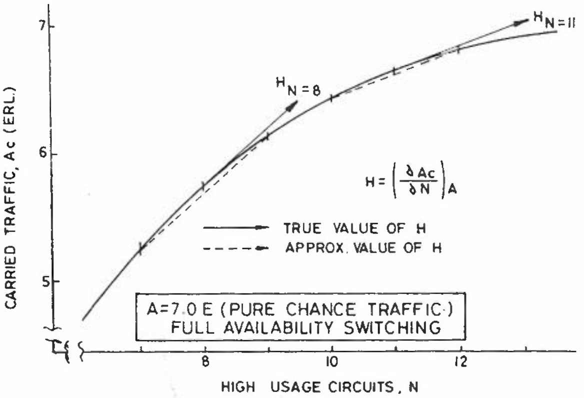


FIG. 4 MARGINAL OCCUPANY APPROXIMATION

Equivalent high usage circuits are then determined using the relationship established earlier in section 2:

$$N_{1eq} = N_1 + \frac{a_1}{H_1}$$

The relationships between N_{eq} and A the pure chance offered traffic for full availability trunking and several values of limited availability trunking with various link losses are illustrated in the graphs of Fig.'s 5 and 6 respectively.

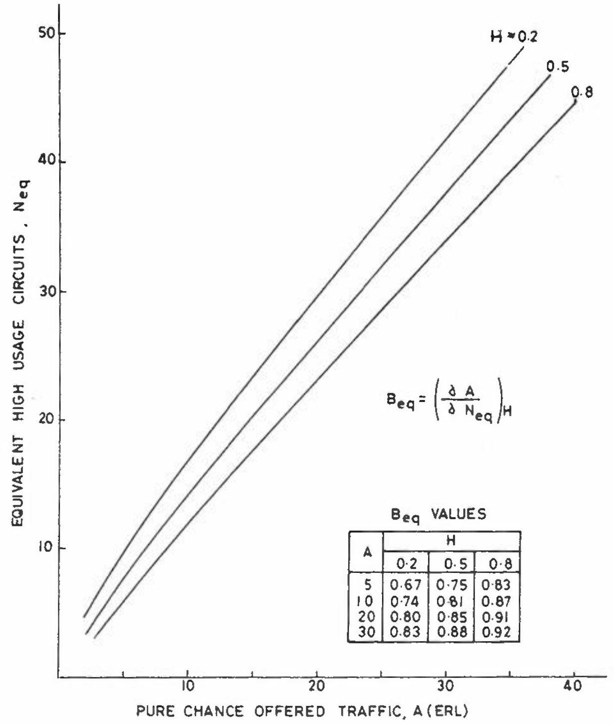


FIG. 5 N_{eq} : OFFERED TRAFFIC (FULL AVAILABILITY)

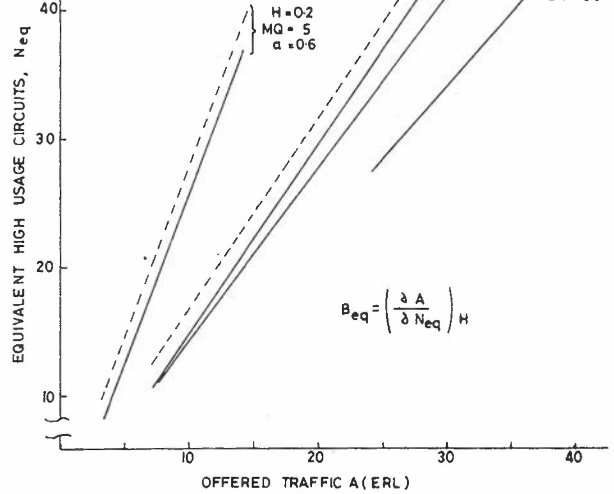


FIG. 6 N_{eq} : OFFERED TRAFFIC (LIMITED AVAILABILITY)

The equivalent marginal capacity B_{eq} - the rate of change of offered traffic with respect to equivalent high usage circuits, H being constant - is given by the inverse of the slope of the curves depicted in Fig.'s 5 and 6. For the full availability case the change in the value of B_{eq} with respect to the change in the offered traffic is not great while for the limited availability examples B_{eq} is markedly constant over the range of traffics shown and these normally apply in practice.

From each data set of nine high usage circuit values sufficient data was available to prepare tables of the relationship between N_{eq} and A (availability, link congestion and H as parameters). For convenience the tables were constructed showing N_{eq} in two components.

(i) That number required for full availability trunking.

(ii) The additional number required because of switching machine inefficiency.

Part of the table of equivalent high usage circuits for marginal occupancy value 0.5 is reproduced in Table 2. It is seen that,

(i) Link rating of 50 for the three stage GV is very similar to the zero link loss graded case,

(ii) the additional number of N_{eq} increases as the probability of link congestion increases and the availability decreases.

TABLE 2 EQUIVALENT HIGH USAGE CIRCUITS, Neq (H=0.5)

A	TOTAL FOR FULL AVAIL	MQ	ADDITIONAL: ZERO LINK CONG	'3' STAGE GV LINK RATING 50	60	70	'2' STAGE GV INLET LOADING 0.5	0.6	0.7	0.8
8	11.75	10	0.01	0.02	0.04	0.12	0.36	0.50	0.68	0.89
		5	0.60	0.64	0.71	0.88	1.26	1.44	1.64	1.86
10	14.26	10	0.05	0.07	0.12	0.24	0.59	0.80	1.04	1.34
		5	1.12	1.18	1.28	1.50	1.99	2.22	2.48	2.76
14	19.14	10	0.53	0.59	0.69	0.91	1.46	1.76	2.11	2.52
		5	2.30	2.40	2.56	2.89	3.60	3.93	4.30	4.70
18	23.90	20	0.01	0.02	0.07	0.19	0.55	0.78	1.06	1.44
		10	1.16	1.25	1.40	1.72	2.49	2.89	3.37	3.92
		5	3.62	3.76	3.97	4.40	5.33	5.77	6.24	6.76
22	28.58	30	0.00	0.00	0.03	0.11	0.37	0.53	0.76	1.07
		20	0.18	0.22	0.29	0.47	0.97	1.27	1.65	2.13
		10	1.87	1.99	2.18	2.60	3.59	4.10	4.69	5.38
		5	5.02	5.19	5.45	5.99	7.14	7.67	8.26	8.89
26	33.20	40	0.00	0.00	0.02	0.07	0.28	0.42	0.62	0.88
		30	0.01	0.02	0.07	0.18	0.53	0.76	1.05	1.45
		20	0.46	0.52	0.62	0.87	1.52	1.90	2.38	2.98
		10	2.66	2.81	3.04	3.55	4.77	5.38	6.10	6.93
30	37.78	40	0.01	0.02	0.04	0.12	0.39	0.57	0.81	1.15
		30	0.06	0.09	0.15	0.31	0.76	1.04	1.41	1.89
		20	0.76	0.84	0.97	1.28	2.09	2.55	3.13	3.85
		10	3.46	3.64	3.92	4.53	5.97	6.69	7.53	8.49
40	49.06	60	0.00	0.01	0.02	0.08	0.30	0.45	0.66	0.95
		40	0.07	0.10	0.16	0.32	0.79	1.07	1.46	1.97
		30	0.52	0.59	0.70	0.96	1.70	2.13	2.69	3.41
		20	1.69	1.82	2.03	2.49	3.70	4.38	5.20	6.23
		10	5.21	5.47	5.83	6.64	8.51	9.45	10.53	11.77
50	60.20	80	0.00	0.01	0.02	0.07	0.26	0.40	0.58	0.85
		60	0.01	0.02	0.06	0.17	0.50	0.72	1.02	1.43
		40	0.39	0.45	0.55	0.79	1.47	1.88	2.42	3.13
		30	1.08	1.18	1.35	1.73	2.75	3.35	4.10	5.07
		20	2.74	2.92	3.20	3.84	5.45	6.34	7.42	8.75
80	93.04	80	0.04	0.07	0.13	0.29	0.76	1.06	1.46	2.02
		60	0.50	0.56	0.68	0.95	1.72	2.19	2.80	3.62
		40	1.71	1.85	2.07	2.58	3.95	4.75	5.77	7.08
		30	3.15	3.37	3.70	4.44	6.38	7.47	8.83	10.54

For the purposes of comparing the effects of different marginal occupancy values Table 3 lists N_{eq} in its two components. The availability has been set = 10.

TABLE 3 EQUIVALENT HIGH USAGE CIRCUITS, Neq (MQ=10)

A	H	TOTAL FOR FULL AVAIL	ADDITIONAL: ZERO LINK CONG	'3' STAGE GV LINK RATING 50	60	70	'2' STAGE GV INLET LOADING 0.5	0.6	0.7	0.8
10	0.2	16.79	0.75	0.84	1.00	1.40	2.50	3.16	3.99	5.06
	0.3	15.81	0.42	0.48	0.59	0.86	1.59	2.02	2.54	3.20
	0.4	14.99	0.21	0.24	0.31	0.50	1.00	1.28	1.62	2.05
	0.6	13.55	0.01	0.02	0.05	0.14	0.37	0.50	0.65	0.82
	0.7	12.83	-	-	0.02	0.06	0.19	0.25	0.33	0.42
	0.8	12.05	-	-	-	-	-	-	-	-
14	0.2	22.06	1.89	2.06	2.33	2.98	4.71	5.70	6.94	8.51
	0.3	20.93	1.27	1.39	1.58	2.02	3.18	3.83	4.63	5.60
	0.4	19.99	0.86	0.94	1.07	1.39	2.20	2.64	3.16	3.80
	0.6	18.32	0.28	0.32	0.38	0.53	0.89	1.08	1.30	1.56
	0.7	17.48	0.07	0.09	0.13	0.21	0.46	0.58	0.72	0.88
	0.8	16.57	0.01	0.02	0.03	0.07	0.18	0.23	0.29	0.35
18	0.2	27.16	3.19	3.44	3.83	4.72	7.08	8.41	10.05	12.12
	0.3	25.90	2.30	2.47	2.74	3.37	4.96	5.83	6.89	8.18
	0.4	24.85	1.68	1.80	1.99	2.44	3.56	4.15	4.86	5.71
	0.6	22.98	0.74	0.80	0.90	1.11	1.63	1.89	2.19	2.53
	0.7	22.05	0.38	0.42	0.48	0.61	0.92	1.07	1.24	1.44
	0.8	21.03	0.08	0.10	0.13	0.19	0.37	0.46	0.55	0.65
22	0.2	32.14	4.65	4.97	5.47	6.62	9.60	11.27	13.32	15.88
	0.3	30.76	3.45	3.67	4.02	4.82	6.85	7.95	9.27	10.87
	0.4	29.62	2.56	2.73	2.99	3.57	5.00	5.75	6.64	7.69
	0.6	27.57	1.29	1.37	1.50	1.79	2.45	2.78	3.16	3.59
	0.7	26.55	0.75	0.81	0.90	1.07	1.48	1.67	1.89	2.14
	0.8	25.43	0.29	0.32	0.37	0.46	0.65	0.74	0.85	0.99
26	0.2	37.04	6.17	6.56	7.17	8.57	12.18	14.18	16.64	19.69
	0.3	35.56	4.62	4.90	5.34	6.33	8.78	10.11	11.69	13.61
	0.4	34.32	3.53	3.74	4.06	4.78	6.51	7.42	8.48	9.75
	0.6	32.12	1.89	1.99	2.15	2.51	3.32	3.72	4.18	4.70
	0.7	31.01	1.20	1.26	1.36	1.58	2.08	2.32	2.58	2.87
	0.8	29.80	0.55	0.59	0.65	0.76	1.02	1.12	1.24	1.38
30	0.2	41.87	-	-	-	-	-	-	-	-
	0.3	40.29	5.88	6.22	6.74	7.90	10.80	12.34	14.19	16.43
	0.4	38.97	4.54	4.79	5.17	6.02	8.06	9.13	10.38	11.85
	0.6	36.62	2.52	2.65	2.85	3.27	4.23	4.70	5.24	5.84
	0.7	35.44	1.67	1.75	1.87	2.14	2.72	3.00	3.31	3.66
	0.8	34.15	0.84	0.89	0.96	1.09	1.38	1.52	1.67	1.82

4. EQUIVALENT MARGINAL CAPACITY VALUES

In the previous Section it was established that the equivalent marginal capacity B_{eq} for full availability trunking does not vary widely with change of offered traffic and for limited availability switching machines is remarkably constant.

In Table 4 values of equivalent marginal capacity, B_{eq} corresponding to the data of Tables 2 and 3 are presented. Although these values are for pure chance offered traffic they can also be used for traffic which is rougher than pure chance, because there is a compensating effect between the rate of change of N and a/H, the two components of N_{eq}, with respect to the offered traffic. The several curves for variance/mean ratio of 2.5 depicted in Fig. 6 illustrate this.

TABLE 4 EQUIVALENT MARGINAL CAPACITY, Beq (A PURE CHANCE)

H	MQ	ZERO LINK CONG	'3' STAGE GV LINK RATING 50	60	70	'2' STAGE GV INLET LOADING 0.5	0.6	0.7	0.8
0.5	40	0.88	0.88	0.87	0.87	0.85	0.84	0.83	0.82
	30	0.86	0.86	0.85	0.84	0.82	0.81	0.80	0.78
	20	0.82	0.82	0.81	0.80	0.78	0.76	0.75	0.73
	10	0.74	0.74	0.73	0.72	0.69	0.68	0.67	0.65
	5	0.66	0.66	0.65	0.64	0.62	0.61	0.60	0.59
0.2	10	0.62	0.62	0.61	0.58	0.54	0.51	0.49	0.46
0.3	10	0.66	0.66	0.65	0.63	0.59	0.57	0.55	0.53
0.4	10	0.70	0.70	0.69	0.68	0.64	0.63	0.61	0.59
0.6	10	0.78	0.78	0.77	0.76	0.74	0.73	0.72	0.71
0.7	10	0.82	0.82	0.81	0.80	0.79	0.78	0.78	0.77
0.8	10	0.86	0.86	0.86	0.86	0.85	0.84	0.84	0.84

5. MARGINAL CAPACITY

In the optimising equations (5) and (9) established in Section 2 marginal capacity values are required for routes 3, 4 and 5 of Fig. 1 where the traffic carried will be rougher than pure chance. The relationship between offered traffic and the number of circuits, for a grade of service of 1 call lost in 200, is shown in Fig.'s 7 and 8.

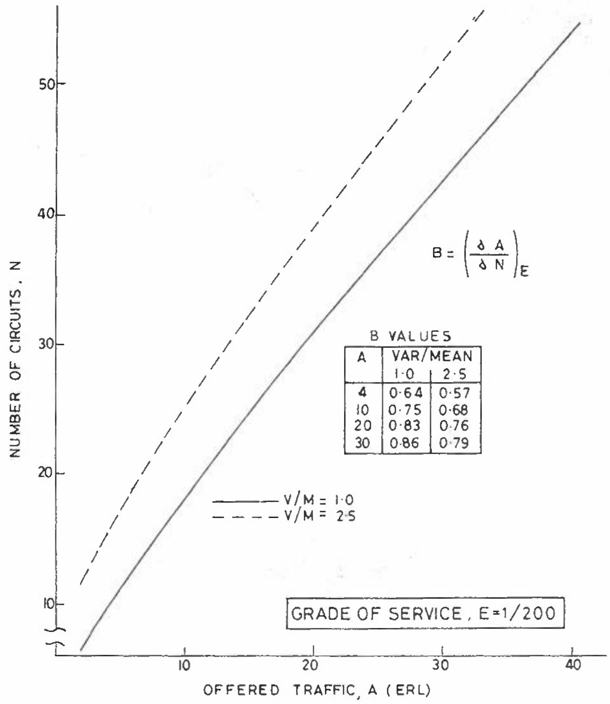


FIG. 7 N: OFFERED TRAFFIC (FULL AVAILABILITY)

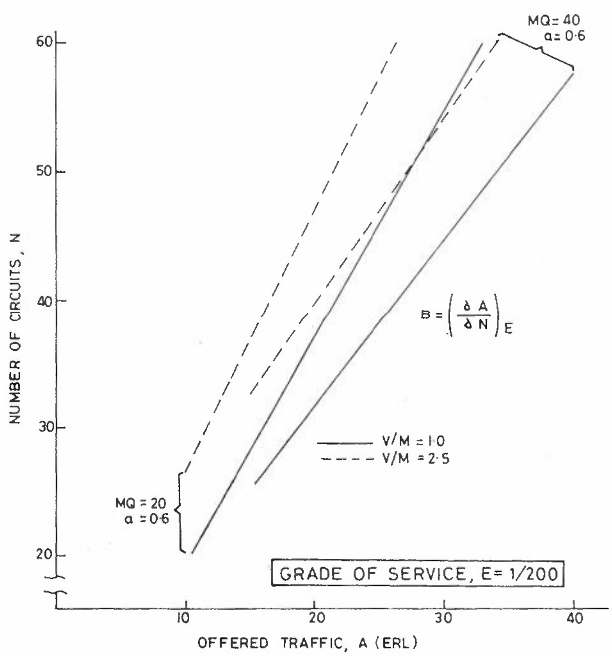


FIG. 8 N: OFFERED TRAFFIC (LIMITED AVAILABILITY)

The curves for both the pure chance traffic and the rough traffic of variance/mean ratio of 2.5 are similar in character to those of Fig.'s 5 and 6, i.e. for the full availability case the change in the value of B with respect to the change in the offered traffic is not great while for the limited availability examples B is markedly constant over the range of traffics shown. The marginal capacity value is affected by the variance/mean ratio of the offered traffic.

In practical designs the marginal capacity value can be estimated more accurately if allowance is made for the roughness of the final route traffics.

Provided consistent estimates of marginal capacity values are made in using equivalent high usage circuits for,

(1) Optimisation of availability,

(2) Comparisons of switching equipment,

(3) Local network analysis,

the results obtained will be quite accurate, i.e. marginal occupancy values are likely to be affected to the same extent.

In Table 5, for pure chance offered traffic, the marginal capacity for a range of availabilities and link congestion are presented. The marginal capacity for rough traffic can be estimated using the approximation,

$$B_x = B_1 - 0.05(x-1)$$

where,

B_x is the marginal capacity at variance to mean ratio of x and
B_1 is the marginal capacity at variance to mean ratio of one.

TABLE 5 MARGINAL CAPACITY, B (A PURE CHANCE)

	GOS = 1/200							
	ZERO LINK	'3' STAGE GV LINK RATING			'2' STAGE GV INLET LOADING			
MQ	CONG	50	60	70	0.5	0.6	0.7	0.8
80	0.91	0.91	0.91	0.91	0.89	0.88	0.87	0.85
60	0.89	0.89	0.89	0.88	0.86	0.85	0.83	0.81
40	0.85	0.85	0.84	0.83	0.80	0.78	0.75	0.72
30	0.80	0.80	0.80	0.78	0.74	0.71	0.67	0.63
20	0.72	0.72	0.71	0.68	0.61	0.57	0.52	0.44

6. EQUIVALENT HIGH USAGE CIRCUITS — DESIGN APPLICATIONS

In the four examples to follow the starting point in each case is to first determine the marginal occupancy of each high usage route. To achieve this it is necessary to estimate marginal capacity values for routes 3, 4 and 5 of Fig. 1, calculate the marginal occupancy of all alternate high usage routes such as route 2 of Fig. 1 and assign an availability parcel to these routes in order to determine the equivalent marginal capacity values. For full availability switching equipment marginal capacity values are determined by estimating the route traffics.

6.1 OPTIMISATION OF AVAILABILITY ALLOCATION

As demonstrated in tables 2 and 3 limited availability switching equipment imposes switching losses in the form of additional or penalty equivalent high usage circuits. For a particular offered traffic value penalty circuits are functions of marginal occupancy, availability and link losses. Optimisation of availability aims at achieving the minimum total cost penalty.

The starting point in the design is to make use of the relationship of equation (11) and determine the cost penalty of each origin/destination traffic parcel assuming all the traffic overflows onto the alternate routes, i.e. the direct routes do not exist.

Cost penalties are then determined assuming that each direct route has been allotted the smallest availability parcel, which for Ericsson ARF 1GV equipment is = 5. From the two sets of cost penalties the cost penalty reduction per unit of availability is determined for each route and a priority queue set up.

When all origin/destination traffic parcels have been processed in this way the smallest availability parcel is assigned to the top value in the queue. The cost penalty reduction in assigning the next availability parcel to this route is then determined and fitted into the queue.

Availability is allocated in this manner to the top value in the queue until it is exhausted. Towards the end of the availability allocation routes at the top of the queue will be passed over if the availability remaining is insufficient to meet the increment required.

The following limited example demonstrates the technique :

Consider the availability allocation to five high usage routes from an origin exchange I. The switching equipment is Ericsson ARF '2' stage GV with an inlet loading of 0.6E. The offered traffics, cost per circuit and marginal occupancy values are as set out in Table 6. The Table shows the penalty equivalent high usage circuits at each value of availability and the cost penalty reduction per unit of availability. The availability is allocated according to the priority numbers established.

TABLE 6 AVAILABILITY ALLOCATION

ROUTE I TO	J	K	L	M	N
COST PER CIRCUIT ($)	200	500	350	490	400
MARGINAL OCCUPANCY	0.2	0.5	0.5	0.7	0.5
OFFERED TRAFFIC (ERL)	14	10	14	14	22
Neq AVAILABILITY,MQ=0	70	20	28	20	44
Neq FULL AVAILABILITY	22.06	14.26	19.14	17.48	28.53
PENALTY Neq MQ= 5	14.68	2.22	3.93	1.25	7.67
" " MQ=10	5.70	0.80	1.76	0.58	4.10
" " MQ=20	1.12		0.46		1.27
" " MQ=30	0.46				0.53
COST PENALTY REDUCTION/UNIT MQ					
MQ = 0 TO 5	1330*	352	345	124	624
MQ = 5 TO 10	359	142	152	66	286
MQ = 10 TO 20	92		46		113
MQ = 20 TO 30	13				30
PRIORITY ORDER FOR MQ ALLOCATION	1 3 11 15	4 8	5 7 13	9 12	2 6 10 14

* 1330 = (70 – (22.06 + 14.68)) x 200 ÷ 5

It will be seen that route IJ although it is the cheapest route warrants an availability allocation first, while route IM is not allocated its first parcel of availability until the other four routes have each had their availabilities increased to 10. The reason for this is that route IM, because of its higher marginal occupancy value, is working at a higher efficiency than the other routes and link congestion has only a minimal effect on the total loss.

6.2 COMPARISON OF SWITCHING EQUIPMENT

In choosing switching equipment it may be necessary to appraise the switching efficiency of the equipment under consideration. Cost penalties in relation to the full availability switching equipment is a convenient method of doing this.

For each type of switching equipment under consideration the availability exercise of the previous section is performed and the total cost penalty determined after all the availability has been allocated.

Such comparisons are necessary when determining if it is economically justified to either instal Ericsson ARF '3' stage GV equipment instead of the '2' stage GV or to convert the latter to '3' stage working.

As the '3' stage costs more per inlet this equipment will be economically justified if the cost difference between the '2' and the '3' stage GV equipment is less than the cost penalty reduction.

6.3 NETWORK ANALYSIS

6.3.1 Tandem Exchanges – Location and Number

In an alternate routed network the number and location of tandem exchanges has a bearing on the cost of switching the network traffic since the more economic the alternate routes of the network are with respect to the high usage routes, the less will be the cost of switching. The alternate routing optimising equations express this in another way :— the more economic the alternate route is with respect to the high usage route the larger will be the value of marginal occupancy of the high usage route.

As the cost of the high usage route is independent of the number and location of tandem exchanges the marginal occupancy of the route can only be influenced by the cost and traffic properties of the alternate routes.

Assuming the costs of switching equipment are independent of location the alternate route costs can only be influenced by the external plant cost components. By locating the 'Y' tandem exchange in the physical path between the origin and destination exchanges and using the same transmission path and cost per unit length as the high usage route the external plant costs of both routes can be made equal. This situation is seldom possible because,

(i) the alternate route transmission standard has to meet the worst case encountered and so the external plant cost components of the alternate route is greater than the external plant cost component of the high usage route, and

(ii) the tandem exchange cannot be located to suit all origin exchanges.

Clearly the greater the number of 'Y' tandems the less the effects of (i) and (ii) above.

Increasing the number of 'Y' tandems on the other hand tends to increase switching costs because,

(i) it increases the number of origin/'Y' tandem routes thereby decreasing the efficiency of these routes, so reducing the value of B_{eq},

(ii) for limited availability switching equipment the availability required for additional 'Y' tandem routes decreases the availability for high usage routes thereby increasing switching costs.

(iii) the number of 'X' to 'Y' tandem routes will be increased and each route will suffer a drop in the value of B.

There are similar advantages and disadvantages in varying the number and location of 'X' tandems.

By parenting destination exchanges on two 'Y' tandems, sited so as to ensure that the majority of origin exchanges have the minimum cost first alternate route path (routes 2 and 5 of FIG. 1), a network cost saving may be achieved. The level of cost reduction resulting from the increase in the marginal occupancy value of route 1 of FIG. 1 can be determined from FIG. 9 where for three values of pure chance offered traffic the number of equivalent high usage circuits relative to the number at marginal occupancy value 0.5 are shown against a range of marginal occupancy values. Both scales are logarithmic and as shown in the graph the grid lines represent 10% increments in the marginal occupancy and 2% increments in

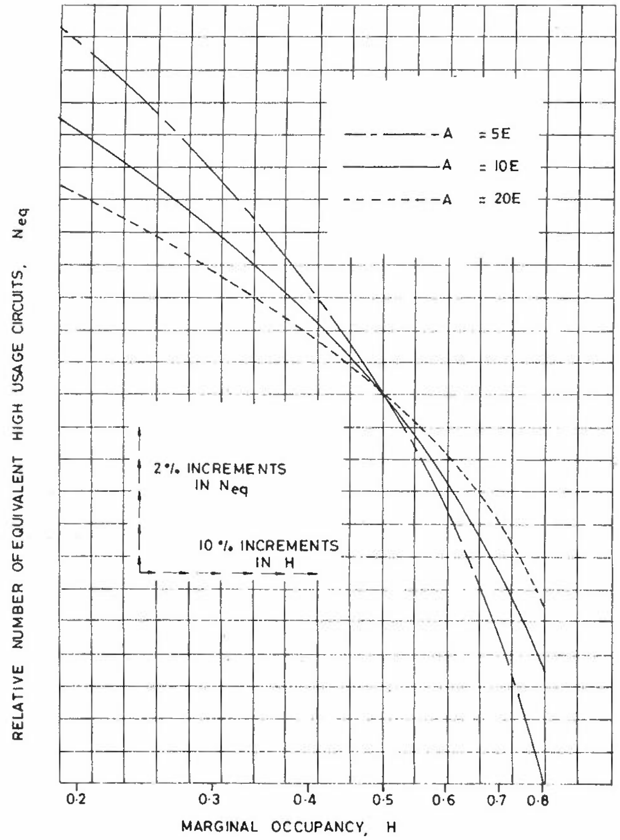


FIG 9 RELATIVE N_{eq} : MARGINAL OCCUPANCY

the relative number of equivalent high usage circuits. The absolute number of equivalent high usage circuits at H = 0.5 for the three traffic values of 5, 10 and 20 ERL. are 7.84, 14.26 and 26.25 respectively.

At H = 0.5 a 10% increase in the value of H would result in a cost saving of,

3.5% for an offered traffic of 5 ERL

2.5% .. .. 10 ERL

2.0% .. .. 20 ERL

The overall network cost saving would be less than these percentages because only about half the origin exchanges would obtain the marginal occupancy improvement and because each terminal exchange is parented on two 'Y' tandems the B value of each final route from the 'Y' tandem to its terminals would be less than what it would be for the one tandem case.

These factors all support what has been determined from numerous tandem studies conducted in Australia and elsewhere — that there is no clearly defined minimum cost solution and that the network costs vary very little as the number and location of tandem exchanges is changed (within limits of course).

6.3.2 Incoming Switching Stages - Design Problems

As exchanges grow in size the limitations of limited availability switching equipment in incoming (I/C) group selector stages becomes evident and the exchange designer is faced with the task of developing the most economic trunking arrangement within the restrictions imposed by the switching equipment and the present method of trunking. Equivalent high usage circuits can be of considerable assistance in examining the possible alternatives.

As shown in Table 7 there are cost penalties incurred if an origin/destination traffic parcel has to be split and this would be necessary if more than one I/C group selector stage is required to handle the incoming traffic.

For example a 10 ERL single route split into two 5 ERL traffic parcels would incur the cost penalties of 2.71, 1.42 and 0.35 equivalent high usage circuits for marginal occupancy values of 0.2, 0.5 and 0.8 respectively. This represents percentage increases of 16.1%, 10.0% and 2.9% respectively in the routing costs.

TABLE 7 Neq - FULL AVAILABILITY TRUNKING

PURE CHANCE TRAFFIC	MARGINAL OCCUPANCY							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
5	10.82	9.75	9.01	8.40	7.84	7.32	6.77	6.20
6	12.37	11.22	10.42	9.76	9.17	8.60	8.02	7.39
8	15.35	14.06	13.16	12.42	11.75	11.10	10.45	9.74
10	18.20	16.79	15.81	14.99	14.26	13.55	12.83	12.05
12	20.98	19.45	18.39	17.51	16.72	15.95	15.17	14.32
14	23.69	22.06	20.93	19.99	19.14	18.32	17.48	16.57
16	26.35	24.63	23.43	22.43	21.53	20.66	19.77	18.81
18	28.97	27.16	25.90	24.85	23.90	22.98	22.05	21.03
20	31.56	29.66	28.34	27.24	26.25	25.29	24.30	23.23
24	36.65	34.68	33.17	31.98	30.90	29.85	28.78	27.62
28	41.65	39.46	37.93	36.65	35.49	34.37	33.23	31.98
32	46.58	44.26	42.64	41.28	40.65	38.86	37.64	36.31
36	51.46	49.01	47.30	45.87	44.57	43.31	42.03	40.62
40	56.29	53.72	51.92	50.43	49.06	47.74	46.39	44.91

At originating exchanges, when traffic routes are split, additional analysis is required and this could be a restriction of the originating group selector stage. If the originating equipment has limited availability additional cost penalties would be incurred above those for full availability trunking because an availability allocation will be required for the additional routes and this will reduce the efficiency of other routes. Split routes would also require greater network management.

Trunking solutions which obviate or limit the splitting of traffic routes will almost certainly be more economic than those based upon splitting traffic routes.

CONCLUSION

The paper introduces the concept of equivalent high usage circuits and shows that the cost of routing an origin/destination traffic parcel can be determined from the cost of the high usage route between origin and destination and a single quantity of equivalent high usage circuits. Equivalent high usage circuits are shown to be functions of the offered traffic, the cost of the high usage route relative to the cost of the alternate routes and the traffic parameters of the switching equipment.

Several examples outlining the uses of the concept have been given and curve fitting equations, relating both actual and equivalent high usage circuits to pure chance offered traffic, are given for full availability switching.

REFERENCES

1. PRATT. C.W., "A Course in Telephone Traffic Engineering", A.P.O. publication prepared by Headquarters Planning Branch, 1967, pp. 75 - 82.

2. BRIDGFORD. J.N., "The Geometric Group Concept and its Application to the Dimensioning of Link Access Systems", Fourth Teletraffic Congress, London, 1964.

OTHER LITERATURE

PRATT. C.W., "The Concept of Marginal Overflow in Alternate Routing", Australian Telecommunication Research, 1967, vol. 1, nos. 1 - 2, pp. 76 - 82.

RAPP. Y., "Planning of a Junction Network in a Multi-exchange Area. I. General Principles", Ericsson Technics vol. 20, no. 1, (1964) pp. 77 - 130.

APPENDIX 1

OTHER ALTERNATE ROUTING PATTERNS AND THEIR OPTIMISING EQUATIONS

The optimising equations for two alternate routing patterns in use in the Australian local network are given without mathematical proof. The simplified sketches alongside the equations will help in understanding the steps involved in deriving the optimising equations.

CASE 1 : EXTENSION OF ALTERNATE ROUTING PATTERN OF FIG. 1.

The routing pattern is an extension of the alternate routing pattern of Fig. 1 of Section 2. The 'X' tandem is allowed high usage routes to selected terminal exchanges as shown in Fig. 10.

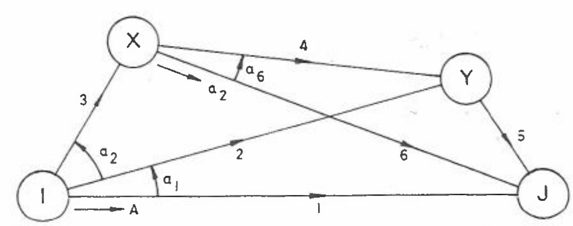
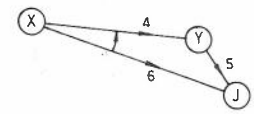
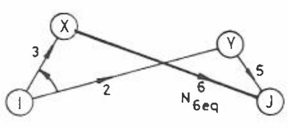


FIG. 10 EXTENSION OF ROUTING PATTERN OF FIG. 1

The optimising equations for the high usage routes XJ, IY and IJ and the routes involved in establishing this equation are as shown :

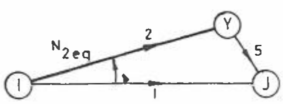


$$\frac{C_6}{H_6} = \frac{C_4}{B_4} + \frac{C_5}{B_5}$$



$$\frac{C_2}{H_2} = \frac{C_3}{B_3} + \frac{C_6}{B_{6eq}} - \frac{C_5}{B_5}$$

$$\text{..} \quad = \frac{C_3}{B_3} + \frac{C_4}{B_4} - \frac{C_6}{1}\left(\frac{1}{H_6} - \frac{1}{B_{6eq}}\right)$$



$$\frac{C_1}{H_1} = \frac{C_2}{B_{2eq}} + \frac{C_5}{B_5}$$

H_2 is dependent on individual values of the H_6 and B_{6eq} parameters and an average value must be used. The average is weighted according to the size of the offered traffic parcel. Where an individual H_2 value is greater than 1.0 (due to the effect of $(1/H_6 - 1/B_{6eq})$ the overflow traffic parcel bypasses route 2 and is offered to route 3.

CASE 2 : SERVICE PROTECTION SWITCHING STAGE

Small origin/destination traffic parcels may not warrant the establishment of high usage routes from the origin exchange. To give these traffic parcels an overall improved grade of service than would result if they were offered directly to the alternate routes of the network a switching stage S is established. The traffic parcel is first offered to the IS route and overflows onto IY. S has high usage routes to terminal exchanges and 'Y' tandems and a final route to an 'X' tandem. Fig. 11 shows the alternate routing pattern. The service protection stage serves a number of origin exchanges and is usually located in the same building as an 'X' tandem.

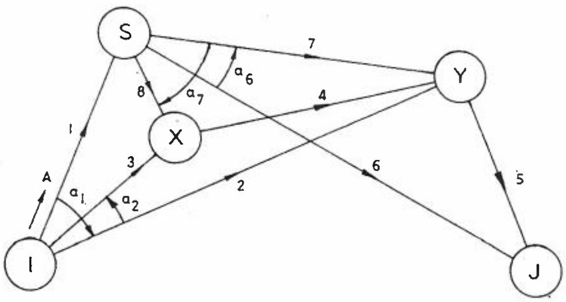
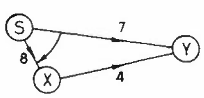
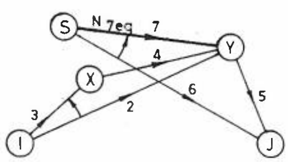


FIG. 11 SERVICE PROTECTION SWITCHING STAGE.

The optimising equations for the high usage routes SY, SJ, IY and IS and the routes involved in establishing these equations are as shown:

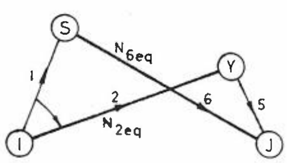


$$\frac{C_7}{H_7} = \frac{C_8}{B_8} + \frac{C_4}{B_4}$$



$$\frac{C_6}{H_6} = \frac{C_7}{B_{7eq}} + \frac{C_5}{B_5}$$

$$\frac{C_2}{H_2} = \frac{C_3}{B_3} + \frac{C_4}{B_4}$$



$$\frac{C_1}{H_1} = \frac{C_2}{B_{2eq}} + \frac{C_5}{B_5} - \frac{C_6}{B_{6eq}}$$

As for case 1 a weighted value of H_1 must be obtained. If H_1 is $\geqslant$ 1.0 (due to the effect of the negative term in the optimising equation) the traffic parcel IJ bypasses route 1 and is offered to route 2.

APPENDIX 2

CURVE FITTING EQUATIONS - FULL AVAILABILITY SWITCHING

1. Equivalent High Usage Circuits - N_{eq}

The relationship between N_{eq} and A the pure chance offered traffic for marginal occupancy values of 0.1 to 0.8 was found to be,

$N_{eq} = P + QA^X + A$
where, P, Q and X are functions of the marginal occupancy H.

$P = -1.259609 + 0.851101\ H^{-0.202437} - 0.564258\ H$

$Q = 1.42843 - 7.54343\ ZS - 1.984073\,(H - 0.55)$
where.

$Z = (\ |H - 0.55|\ )^{4.155342}$
$S = +1$ for $H \geqslant 0.55$ and -1 for $H < 0.55$

$X = 0.495195 + 0.007328\ H^{0.49} - 0.002438\ H$

Between the limits of H = 0.1 and 0.8,

P ranges from 0.040460 to −0.820586
Q " " 2.594510 to 0.908659
X " " 0.497322 to 0.499814

The equivalent marginal capacity can be determined from,

$1/B_{eq} = XQA^{X-1} + 1.0$

$\doteq 0.5\ QA^{-0.5} + 1.0$

2. Actual high Usage Circuits - N

The relationship between N and A the pure chance offered traffic for marginal occupancy values of 0.1 to 0.8 was found to be,

$N = R + TA^{0.5} + A$
where R and T are functions of the marginal occupancy H.

$R = 0.527355 - 2.233955\ H^{0.5} - 0.06694\ H$

$T = 0.923485 - 12.59178\ ZS - 3.396178\,(H - 0.3915)$
where,

$Z = (\ |H - 0.3915|\ )^{3.543912}$
$S = +1$ for $H \geqslant 0.3915$ and -1 for $H < 0.3915$

Between the limits of H = 0.1 and 0.8

R ranges from −0.185778 to −1.524307
T " " 2.072990 to −0.991309

3. Actual Circuits for Grade of Service Criteria

For grades of service 1 in 200 and 1 in 500 the relationship between N and A is as follows:-

$N_{1/200} = 0.745148 + 2.971970\ A^{0.428036} + 0.985860\ A$

$N_{1/500} = 0.990524 + 3.260262\ A^{0.441454} + 0.990887\ A$

Discussion

L.T.M. BERRY, Australia : *Have the assumptions of the model been subjected to validation? For example (i) the equation* $N_3 \underline{n} \frac{a_2}{B_3}$ *assumes* a_2 *is small. It has been found that for certain OD pairs a minimum cost solution results in more traffic being carried on 3rd choice routes than 2nd. i.e.* a_2 *may be reasonably large. (ii) An assumption implicit in the model is that if the cost of routing traffic for any one OD pair can be decreased the solution is not optimal. It has been shown by Harris that a "User optimal" solution is significantly more expensive for a Metropolitan network than a "System optimal" solution.*

J.S. HARRINGTON, Australia : *The assumptions of the model have been tested in the Perth, Australia, metropolitan network where the costs agreed fairly closely with those obtained from the proper dimensioning of the network. The difference in costs could be attributed to the constant number of circuits on the IX, XY and YJ routes (see figs. 7 & 8 where for VAR/MEAN > 1 the relationship between offered traffic A and N of circuits is,*

$$N = \text{Const} + f(A) \quad)$$

I must point out that my method is not a dimensioning method but rather a tool to aid dimensioning and to understand the factors that affect network design. This means that the cost difference obtained above is likely to be fairly constant for a number of network configurations. The approximation $N_3 \approx a_2/B_3$ *holds more for limited availability switching than for the full availability case but the value of* B_3 *so obtained (see figs. 7 & 8) is fairly constant over the range of* a_2 *and after all it is* B_3 *that is required.*

I would agree that for the full availability switching machine the first alternate route, IY, could well be by-passed in many cases because,

(i) the traffic overflowing from the H.U. route, IJ of fig. 1 will be less, and

(ii) the efficiency of the final routes, IX, XY and YJ greater,

than the corresponding values for the limited availability switching machine.

W. LÖRCHER, Germany : *In Chapter 2 of your paper you make an assumption, denoted as result No. 2 (ii)* $N_3=a_2/B_3$ *and* $N_4=a_2/B_4$*, where it is stated that the group sizes of final routes are given by the quotient of offered traffic and the marginal capacity.*

Is this assumption only applied for the optimization of high usage trunk groups or also for dimensioning of the final trunk group itself.

J.S. HARRINGTON, Australia : *The assumption is not used to dimension final routes but to obtain the value of the marginal occupancy of the high usage junction routes.*

W. LÖRCHER, Germany : *In Chapter 5 you give an approximation formula for the marginal capacity of trunk groups with offered overflow traffic. This marginal capacity depends on the marginal capacity of a trunk group with offered PCT1 and the variance to mean ratio of the offered overflow traffic. Can you give an impression on the accuracy of this formula.*

J.S. HARRINGTON, Australia : *The formula has only been checked in a few cases and so I cannot guarantee its accuracy. However, as the value of* B_3*,* B_4 *and* B_5 *have only a minor effect on the value of marginal occupancy of the H.U. route I do not consider a slight error in the estimation of B to be a limitation to the method which is a network analysis and dimensioning tool and not a dimensioning process.*

BIOGRAPHY

JAMES S. HARRINGTON entered the PMG Department in 1947 as Technician-in-Training. In 1956 he was promoted to Trainee Engineer from Supervising Technician and in 1959 graduated as Engineer.

From 1959 to 1972 he worked in the Planning & Programming Branch in Perth, having special interest in the switching design of Metropolitan Networks.

He served two short term missions for the I.T.U. acting as Telephone Traffic Adviser to CANTV (Venezuela) in 1967/68 and Telephone Traffic and Internal Plant Planning Adviser to STB (Singapore) in 1972/73.

After 3½ years in the Metropolitan Operations Area he returned to the Planning & Programming Branch in October 1976 and is currently Supervising Engineer, Traffic Engineering Section.

Dimensioning of Alternative Routing Networks Offered Smooth Traffic

J. RUBAS

Telecom Australia, Melbourne, Australia

ABSTRACT

This paper reviews the methods of dimensioning networks employing alternative routing and discusses the use of Binomial distribution model when the offered traffic is smoother than pure chance. Instead of the Poisson-based "equivalent random" method the more accurate direct computation of overflow traffic moments is proposed. A comparison is made between this and two other methods for accuracy and convenience of computation. Sample dimensioning graphs for full and limited availability trunk groups are appended.

1. INTRODUCTION

During the last twenty years the technique of alternative routing has become firmly established in most large communication networks. A number of different methods have been developed to dimension these networks. Virtually all of these methods are based on the assumption of Poisson-distributed traffic being offered to first choice routes.

Alternative routing can be profitably used also in situations where traffic is generated by a limited number of sources and is, therefore, smoother than pure chance. These situations arise in private automatic branch exchanges, small rural automatic exchanges, and subscriber switching stages of public exchanges. Classical, Poisson-based methods are unsuitable for dimensioning such systems, as they tend to overestimate circuit requirements. It has been shown that the Binomial distribution model gives a better representation of the traffic offered in the above cases, particularly when all traffic sources have similar calling intensities. The manipulation of this model, however, presented considerable mathematical difficulties and so far its application has been largely confined to the dimensioning of constant grade of service routes without overflow, using tables or graphs based on the Engset Loss Formula (e.g. Ref. 1).

In order to use this model for dimensioning alternative routing systems offered traffic from a limited number of sources a method is required to compute or estimate at least the first two cumulants of the overflow traffic. The first viable method, based on charts produced from simulation results, was published in a little-known paper by Wormald in 1968 (Ref. 2). In 1973 Bretschneider published an iterative numerical method (Ref. 3), which permitted extension of Wilkinson's equivalent random technique to smoother than pure chance traffics. In the same year Schehrer (Ref. 4) and Henderson (Ref. 5) independently obtained an exact analytical solution to the problem of computing the first two cumulants of traffic overflowing a full availability group of trunks offered Binomially-distributed traffic. Finally, the following year Harris and Rubas (Ref. 6) described an accurate general method for calculating the overflow moments in full and limited availability trunk groups. This opened the way to accurate dimensioning of all types of trunking schemes offered smoother than Poisson-distributed traffic.

2. REVIEW OF ALTERNATIVE ROUTING

The objective of alternative routing is to carry traffic at minimum cost for a given grade of service. In the simplest case of one direct and one alternative route the optimising equation has the following form:

$$C_d/H = C_o/q \qquad \text{........(1)}$$

Here C_d and C_o are costs per circuit of the direct and the alternative route. H is the marginal occupancy of the direct route, defined by

$$H = \left(\frac{\partial Y_1}{\partial N_1}\right)_{A_1} \qquad \text{........(2)}$$

where A_1 is the traffic offered and Y_1 the traffic carried by N_1 direct route circuits. q is the marginal capacity of the alternative (overflow) route, defined by

$$q = \left(\frac{\partial A_2}{\partial N_2}\right)_B \qquad \text{........(3)}$$

where A_2 is the traffic offered to the overflow route of N_2 circuits, which is to be dimensioned for a congestion probability B.

If several alternative routes are possible, it is necessary to set up an optimising equation for each pair of alternatives. In such cases the accuracy of optimisation calculation is improved by employing another derivative, g, which is the marginal increase in overflow, a, per unit increase in the traffic offered, A, while the number of circuits, N, in the route under consideration is kept constant. Expressed mathematically,

$$g = \left(\frac{\partial a}{\partial A}\right)_N \qquad \text{........(4)}$$

The first step in the dimensioning process is to obtain the economical marginal occupancies, H, for the first choice routes from appropriate optimising equations and then compute the number of circuits in each first choice route that will give the required marginal occupancy. The next step is to compute mean and variance of all traffics overflowing from first choice routes and to determine the economic marginal occupancies of the second choice routes. Mean and variance of traffics offered to second choice routes are obtained by adding the respective moments of all traffic parcels offered to particular second choice routes, which are then dimensioned to give the previously computed marginal occupancies.

The above process is repeated until all choices for high usage routing have been exhausted and we are left with only one choice - the final route. This route is normally dimensioned for a specified loss probability, which is chosen to ensure that a satisfactory overall grade of service is provided.

The dimensioning work can be done manually, with the aid of graphs or tables, or by means of suitable computer programs. Although manual calculations have been largely replaced by computers, there still is a place for graphical aids in spot checks of individual routes and simple networks. In cases where the dimensioning algorithm involves very complex mathematical models, graphical methods may even be more efficient than numerical ones.

3. APPLICATION OF BINOMIAL MODEL

The natural application of the Binomial model is in situations where the traffic is generated by a limited number of sources, each having the same calling intensity. If there is no congestion, the probability of finding x out of S sources busy is

$$P(x) = \binom{S}{x} \cdot a^x \cdot (1 - a)^{S-x} \qquad \ldots\ldots\ldots(5)$$

where a = A/S is the average traffic offered per source.

The mean of the above distribution, A, is, therefore,

$$A = S.a \qquad \ldots\ldots\ldots(6)$$

and its variance, V, is

$$V = A(1 - a) \qquad \ldots\ldots\ldots(7)$$

Replacing a by A/S and solving for S gives the convenient relationship between A, S, and V :-

$$S = A^2/(A - V) \qquad \ldots\ldots\ldots(8)$$

To use this model for the dimensioning of high usage and final choice routes we must be able to compute mean and variance of overflow traffics and also the marginal occupancy, marginal capacity, and, possibly, marginal overflow for any number of circuits offered Binomially distributed traffic. As an approximation, the Binomial model can also be used to represent other types of smooth traffic (A > V), e.g. truncated Poisson (Erlang). It must be realised, however, that with the Binomial model the magnitudes of overflow moments depend on whether the overflow traffic is lost, or carried on another rocte. For example, traffic lost from a fully available group of N trunks offered A erlangs by S independent traffic sources will be given by A.B(A, N, S), where B(A, N, S) is the Engset Loss probability; if the overflow traffic is carried on another route, however, its magnitude will be less than A.B(A, N, S).

It has been shown (Refs. 4, 5) that for full availability trunk groups offered Binomial traffic the cumulants of the overflow can be computed using Wallström's method (Ref. 9) for state-dependent call intensity functions. The Binomial call intensity function has the form

$$A(x) = \alpha.(S - x) \qquad \ldots\ldots\ldots(9)$$

where α is the traffic offered per free source. The Binomial distribution admits only S+1 states, hence Wallström's infinite sums have to be replaced by finite ones.

Where access to direct and alternative routes is limited by switch outlet availability or link congestion, other methods have to be used to compute mean and variance of the overflow traffic. Since there is a finite number of states, the problem can be tackled by setting up a system of linear equilibrium state equations, which can be solved for state probabilities P(n, m) by matrix methods (Ref. 6). Having computed state probabilities, the required moments of overflow traffic are obtained from standard statistical formulae:-

$$\text{mean} = \sum_{n=0}^{N} \sum_{m=0}^{S-n} m \cdot P(n,m) \qquad \ldots\ldots\ldots(10)$$

$$\text{variance} = \sum_{n=0}^{N} \sum_{m=0}^{S-n} m^2 \cdot P(n,m) - (\text{mean})^2 \qquad \ldots\ldots(11)$$

The restricted access to route can be defined by the distribution of conditional blocking probabilities, which are determined from the configuration and average traffic loading of the switching stage from which the route is trunked.

For manual dimensioning, graphs of mean and variance of overflow have been constructed on the assumption that access to direct routes can be defined by a Geometric distribution of blocking probabilities, which can be defined by a single parameter, p. This "geometric group" model has been found very useful in dimensioning routes accessed through link-trunked switching stages. By definition, p is the probability of the next call being blocked when there is only one free circuit in the traffic route under consideration. Then the probability of blocking with two free circuits is p^2, with three free circuits it is p^3, and so on. The geometric group model was first proposed by Smith in 1961 (Ref. 10) and is still in use.

To dimension the final choice route we must know at least the first two cumulants (mean and variance) of the traffic offered to it and the specified grade of service. The cumulants are computed by adding the means and the variances of all traffics overflowing to the route in question. If the mean, A, of the aggregate traffic is greater than its variance, V, the Binomial distribution can be used to approximate it. After computing the equivalent number of sources from equation (8), the required number of trunks can be estimated from an appropriate traffic capacity table or graph based on the Engset Loss formula (Refs. 1, 11, 12) if full availability access to the final route is provided. No traffic capacity graphs or tables have been published for dimensioning limited availability trunk groups offered Binomially distributed traffic. If the restricted access can be defined by a conditional blocking probability distribution (e.g. the Geometric distribution), the required number of trunks, N can be computed iteratively from the following equation for call congestion:

$$B = \frac{\alpha}{A} \sum_{x=0}^{N} (S-x).b(x).P(x) \qquad \ldots\ldots\ldots(12)$$

where A is the mean offered traffic in erlangs, b(x) is the blocking probability in state x, S is the equivalent number of sources, P(x) is the probability that x circuits are occupied and α is the traffic per free source, given by

$$\alpha = A/(S - A.(1 - B)) \qquad \ldots\ldots\ldots(13)$$

The state probabilities can be computed recursively from the following equations :

$$P(x+1) = \frac{\alpha.(S-x).(1-b(x))}{x+1} \cdot P(x) \qquad \ldots\ldots(14)$$

$$\sum_{x=0}^{N} P(x) = 1 \qquad \ldots\ldots\ldots(15)$$

Time congestion for the limited availability route is given by the sum

$$E = \sum_{x=0}^{N} b(x).P(x) \qquad \ldots\ldots\ldots(16)$$

Equation (16) can be shown to be equivalent to

$$E = \frac{\binom{S}{N}\alpha^N \prod_{i=0}^{N-1} (1 - b(i))}{1 + \sum_{x=1}^{N} \{\binom{S}{N}\alpha^x . \prod_{i=0}^{x-1} (1-b(i))\}} \qquad \ldots\ldots\ldots(17)$$

Since A and V of the traffic offered to the route are known, S can be computed from equation (8).

The blocking coefficients b(x) depend on the configuration and loading of the switching stage giving access to the route. As stated before, the access system blocking can be represented by a geometric series or another suitable mathematical model.

If the mean of the aggregate overflow traffic is less than its variance, the traffic can be represented by the Negative Binomial distribution, or the dimensioning can be completed using Wilkinson's "equivalent random" technique. The latter method is very well known and requires no elaboration; the use of Negative Binomial model will be briefly reviewed here.

The probability density function of the Negative Binomial distribution is defined by

$$P(x) = \binom{S+x-1}{x} . a^x.(1-a)^S \qquad \ldots\ldots\ldots(18)$$

$$0 \leq a \leq 1 \qquad S \geq 0$$

The parameters a and S can be determined from the mean (A) and variance (V) of the offered traffic distribution:

$$a = 1 - A/V \qquad \ldots\ldots\ldots(19)$$

$$S = A^2/(V-A) \qquad \ldots\ldots\ldots(20)$$

Call congestion for a fully available group of N trunks offered Negative Binomial traffic is given by

$$B = \frac{\alpha^N \binom{S+N}{N}}{\sum_{i=0}^{N} \binom{S+i}{i} \alpha^i} \qquad \ldots\ldots\ldots(21)$$

where

$$\alpha = A/(S + A . (1 - B)) \qquad \ldots\ldots\ldots(22)$$

For a limited availability group, where access is defined by blocking coefficients b(x), call congestion can be computed from the following sum :

$$B = (\alpha/A) \sum_{x=0}^{N} (S+x).b(x).P(x) \ldots\ldots\ldots(23)$$

The parameter α is again defined by (22) and the state probabilities P(x) can be computed from equations (14) and (15), with (S-x) in eqn. (14) replaced by (S+x). Time congestion can be computed from the same general equation (16), which applies to all traffic distributions.

Where full availability conditions exist, call congestion for both Negative and Positive Binomial distributions (truncated at N trunks) can be more easily computed from the following recurrence, starting with B(0)=1:

$$B(N) = \frac{\alpha.(S \pm N).B(N-1)}{N + \alpha.(S \pm N).B(N-1)} \qquad \ldots\ldots\ldots(24)$$

In the above equation + sign between S and N applied to Negative and - sign to Positive Binomial. Parameters α and S are as defined previously for the two distributions. Note that S, the equivalent number of sources does not have to be an integer, which facilitates the fitting of Binomial models to other traffic distributions; it will be appreciated that for non-Binomial distributions equations (8) and (20) will not, generally, give integral values for S.

The above recurrence and associated equations (13) and (22) have been programmed for the HP65 pocket calculator and provide a convenient means of computing congestion for either distribution.

4. COMPARISON WITH OTHER METHODS

Binomial distribution is the best model for traffic generated by a limited number of sources with similar calling intensities. However, a fair amount of calculation effort is required to compute the marginal occupancy and the moments of overflow traffic, particularly for limited availability routes. In the latter case accurate calculation of overflow moments each time they are required is impracticable. For this reason a set of accurately plotted graphs have been produced for the mean and variance of overflow traffic and the marginal occupancy ("improvement factor"). Samples of these graphs are given in the Appendix.

Obviously, it would be an advantage to have simpler calculation routines, which could be incorporated in computer programs for dimensioning of direct and alternative routes. In the following tables two approximate methods are compared with the Binomial model in the computation of overflow traffic mean and variance when small trunk groups are offered traffic from a limited number of sources. The first approximation assumes infinite number of traffic sources (Poisson input) while the second employs extended equivalent random technique described in Ref. 3. Tables 1a and 1b give moments of traffic overflowing from a full availability group of trunks, while Tables 2a and 2b compare overflows from a limited availability group, access to which is defined by geometrically distributed conditional blocking probabilities; the parameter p is numerically equal to the probability of blocking when only one free circuit remains in the trunk group.

5 Sources, Full Availability

Traffic (erl.)	No. of Trunks	Mean and Variance of Overflow			
			Binomial	Poisson	ERM (Ref. 3)
3.0	1	m	2.166	2.250	2.118
		v	1.101	2.587	1.167
"	2	m	1.431	1.588	1.324
		v	0.772	2.066	1.015
"	3	m	0.767	1.038	0.692
		v	0.487	1.488	0.679
"	4	m	0.254	0.618	0.278
		v	0.189	0.945	0.317
"	5	m	0.000	0.330	0.082
		v	0.000	0.519	0.098

Table 1a

Comparison of Overflow Traffic Moments, Full Availability Access to Route

10 Sources, Full Availability

Traffic (erl.)	No. of Trunks		Mean and Variance of Overflow		
			Binomial	Poisson	ERM (Ref. 3)
1.0	5	m v	0.001 0.001	0.003 0.004	0.002 0.002
6.0	6	m v	1.226 0.946	1.590 2.746	1.097 1.293
"	7	m v	0.665 0.575	1.110 2.019	0.615 0.808
"	8	m v	0.257 0.242	0.731 1.372	0.296 0.411
"	9	m v	0.048 0.046	0.451 0.855	0.120 0.168

Table 1b

Comparison of Overflow Traffic Moments,
Full Availability Access to Route

5 Sources, Limited Availability, p = 0.3

Traffic (erl.)	No. of Trunks		Mean & Variance of Overflow		
			Binomial	Poisson	ERM (Ref. 3)
3.0	1	m v	2.260 1.020	2.323 2.558	2.118 1.167
"	2	m v	1.579 0.816	1.717 2.060	1.324 1.015
"	3	m v	0.984 0.592	1.199 1.542	0.692 0.679
"	4	m v	0.511 0.367	0.782 1.053	0.278 0.317

Table 2a

10 Sources, Limited Availability, p = 0.4

Traffic (erl.)	No. of Trunks		Mean & Variance of Overflow		
			Binomial	Poisson	ERM (Ref. 3)
3.0	1	m v	2.333 1.786	2.357 2.553	2.190 1.866
"	2	m v	1.729 1.442	1.781 2.069	1.469 1.519
"	3	m v	1.206 1.083	1.284 1.577	0.882 1.058
"	4	m v	0.778 0.741	0.875 1.116	0.458 0.604
"	5	m v	0.456 0.452	0.560 0.725	0.200 0.275
"	6	m v	0.242 0.243	0.334 0.430	0.073 0.099

Table 2b

Comparison of Overflow Traffic Moments,
Limited Availability Access to Route

As can be seen from the above tables, the Poisson model consistently overestimates both the mean and the variance of overflow traffic. The "equivalent random" method gives a fairly good estimate of both overflow moments under full availability conditions (except when N approaches S); the error, naturally, increases when availability is restricted. Both approximate models show significant overflows at N=S, when there should be no overflow.

The calculation time is shortest for the Poisson model and longest for the Binomial; the equivalent random method is intermediate between the other two, but would, probably, be still too slow for large scale computations involving larger trunk groups. It appears likely that curve fitting of accurately computed overflow mean and variance graphs will yield more efficient computation routines, but this has not yet been investigated. In the interim, manual calculation with the help of graphs (Ref. 14) may prove to be more economical, particularly where full availability conditions are not provided.

5. CONCLUSIONS

This paper has shown how the Binomial distribution model can be used to improve the accuracy of dimensioning alternative routing networks offered traffic from a limited number of sources. The model is suitable for dimensioning both high usage and final choice routes, reached with full or restricted availability. The improvement in accuracy is obtained through increased computing effort.

6. ACKNOWLEDGEMENT

The author wishes to thank Dr. Richard Harris who developed the computer programs for the calculation of overflow traffic moments and marginal occupancies for restricted availability trunk groups offered Binomially distributed traffics. Thanks are also due to Edward Heim who prepared the program for the computation of overflow mean and variance using the extended equivalent random method.

7. REFERENCES

(1) Driksna, V.V. and Wormald, E.G. "Engset Traffic Tables". The Telecommunication Journal of Australia; June, 1966.

(2) Wormald, E.G. "Consecutive Overflow Telephone Traffic Graphs". Electrical Engineering Transactions, Institution of Engineers, Australia; September, 1968.

(3) Bretschneider, G. "Extension of the Equivalent Random Method to Smooth Traffics" Proceedings of 7th I.T.C., Vol. 2, 1973.

(4) Schehrer, R. "On the Calculation of Overflow Systems with Full Available Groups and Finite Source Traffic". 10th Report on Studies in Congestion Theory, University of Stuttgart, 1973.

(5) Henderson, J. Private Communication, 1973.

(6) Harris, R.J. and Rubas, J. "Calculation of Overflow Traffic Moments for Full and Limited Availability Systems with Binomially Distributed Offered Traffic". Australian Telecommunication Research, Vol. 8, No.2, 1974.

(7) Wilkinson, R.I. "Theories for Toll Traffic Engineering in the U.S.A." Bell System Technical Journal, March, 1956.

(8) Pratt, C.W. "The Concept of Marginal Overflow in Alternate Routing". Australian Telecommunication Research, Vol. 1, Nos. 1 & 2, 1967.

(9) Wallstrom, B. "Congestion Studies in Telephone Systems with Overflow Facilities". Ericsson Technics, No.3, 1966.

(10) Smith, N.M.H. "More Accurate Calculation of Overflow Traffic from Crossbar Group Selectors". 3rd I.T.C. Doc. No.36, 1961.

(11) Rubas, J. "Table of the Engset Loss Formula". Traffic Engineering Report No.11, Postmaster-General's Department, Melbourne, 1969.

(12) Agosthazi, M., Gosztony, G., and Uxa, E. "Engset Charts for the Design of Loss Systems with a Finite Number of Traffic Sources and Full Availability". Budavox Telecommunication Review, 1973.

(13) Mina, R.R. "Some Practical Applications of Teletraffic Theory". Proceedings of 5th I.T.C., 1967.

(14) Harris, R.J. and Rubas, J. "Overflow Traffic Graphs for Smooth Traffic." Telecom Australia, 1976.

APPENDIX

This appendix contains samples of graphs prepared for manual dimensioning of alternative routing systems offered smooth traffic. Figures 1, 2, and 3 apply to fully available trunk groups, while Figs. 4, 5 and 6 are drawn for trunk groups reached through a grading or a link-trunked switching stage. In the latter case the blocking coefficients are assumed to be geometrically distributed with parameter p, which represents the blocking probability with only one free trunk in the route.

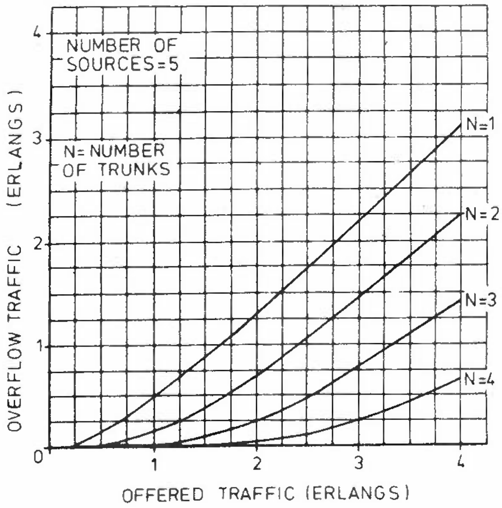


Figure 2

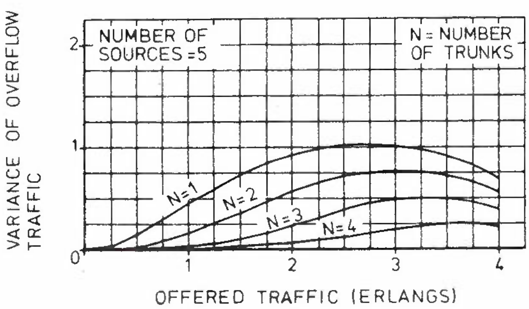


Figure 3

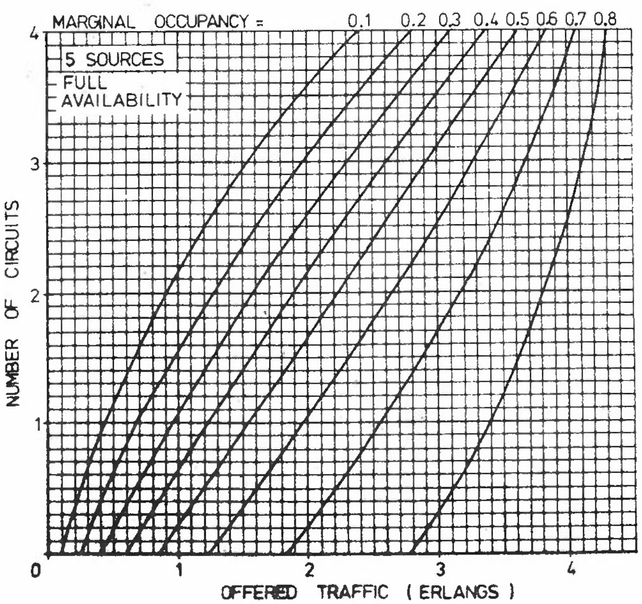


Figure 1

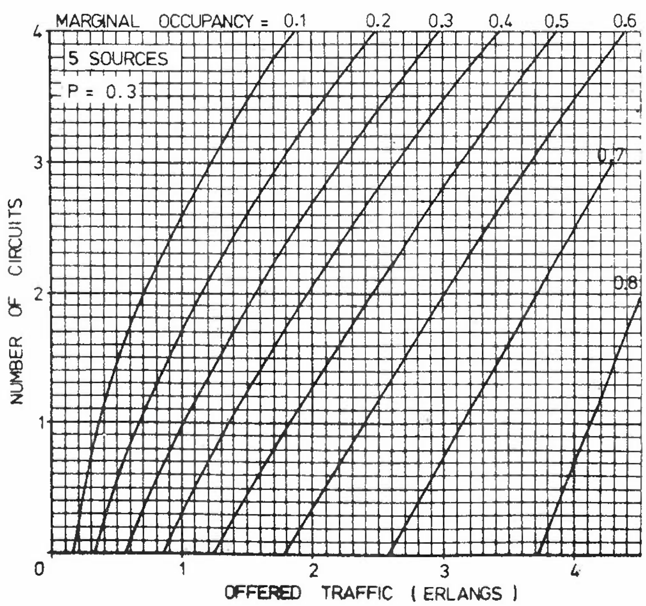


Figure 4

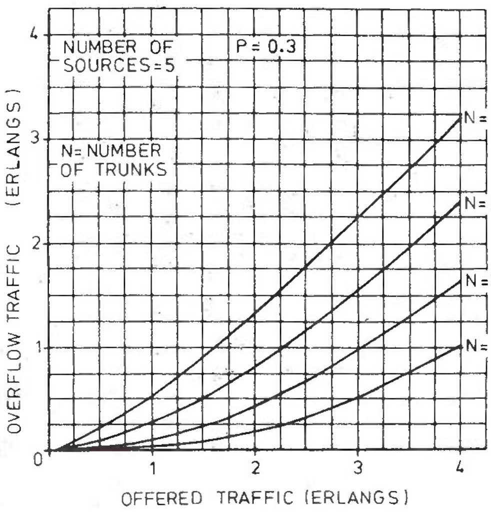


Figure 5

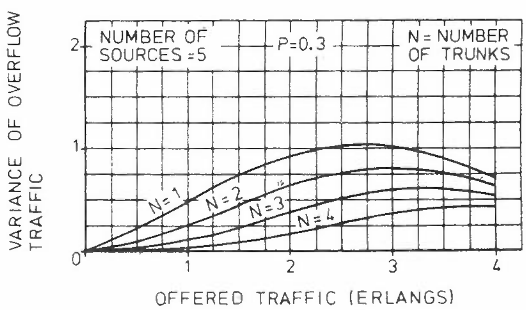


Figure 6

Discussion

J. DE BOER, Netherlands : *First, personally I would prefer the term "Engset traffic" instead of "binomial traffic" because "binomial" is usually associated with the situation where the number of sources is equal to the number of lines. Secondly, consider the situation of two groups of lines of say n_1 and n_2 lines respectively. The aggregate of the two carried traffics is offered to a third group of lines. Then one can imagine that the fictitious number of sources c_f is smaller than $n_1 + n_2$, implying that no more than c_f lines can be occupied at a given moment, in contradiction with the fact that $n_1 + n_2$ can be occupied. Does this possibility really occur sometimes.*

J. RUBAS, Australia : *In an alternative routing network most of the calls overflowing from high usage routes are carried on later choice routes. As no truncation of the offered traffic distribution occurs at this stage the Binomial model is more appropriate than the Engset one; the reverse is true when we are dimensioning routes without overflow facilities, as stated in my paper.*

In answer to your second question I must agree that it would be possible to get an equivalent number of sources c_f, which is less than the combined number of trunks $n_1 + n_2$ postulated in your hypothetical example. But this would only occur if one or both of the two trunk groups were grossly overdimensioned and carried traffic with no loss (e.g. if they were offered traffic by a small number of real sources, the sum of which was less than $n_1 + n_2$).

J. MATTES, Australia : *Where full availability conditions exist, call congestion for both Negative and Positive Binomial distributions can be computed from relation (24) which gives B(N) in terms of B(N-1) and α. However, according to (22), α is a function of B.*

Could the author make it clear whether α in (24) is α(Bn) or α(Bn-1). (i.e. is iteration required at each stage of the recursion.)

J. RUBAS, Australia : *In equation (24) α is a function of B(N) for either distribution; for the Binomial distribution it is defined by equation (13) and for the Negative Binomial distribution by equation (22). Because α is a function of congestion, iteration is necessary to evaluate B(N).*

R. SCHEHRER, Germany : *As far as I see, there exist 2 different types of smooth traffic: First, smooth traffic which has its origin in Poisson or Engset traffic, respectively, and has been smoothed due to the loss in one or more preceding trunk groups, and the second definition, applied in your paper, according to which all traffics are considered to be smooth as soon as their variance-to-mean ratio is less than one. This definition would include the well-known pure chance traffic of second kind (PCT2) which is random traffic from a finite number of traffic sources. It seems to me that these 2 types of so-called smooth traffic should be distinguished. How do you think about this.*

J. RUBAS, Australia : *My definition of smooth traffic is very general and includes all traffic distributions for which the mean is greater than the variance. I agree with Dr. Schehrer that smooth traffics in accordance with my definition could be further subdivided into traffics generated by limited number of sources (referred to as PCT2 traffic) and the smoothed traffics resulting from truncation of peaks by preceding trunking, regardless of the type of the original distribution. In this context I would like to note that while the Binomial distribution is a good model for the limited source traffic, it is less accurate in representing other kinds of smoother than pure chance traffic.*

D.J. SONGHURST, U.K. : *When the binomial distribution model is used for traffic generated by a limited number of sources, and the first two overflow moments are used, how accurately can final routes be dimensioned.*

J. RUBAS, Australia : *If all traffic sources have the same calling intensity and originate calls quite independently of each other, the first two moments of the Binomial distribution are quite sufficient for accurate dimensioning of both final choice and high usage routes, as long as the variance to mean ratio of the traffic offered to these routes is less than one. When the overflow traffic degenerates to the point when variance exceeds the mean, one must either use another distribution model (e.g. Negative Binomial), or fall back on the equivalent random technique, which may introduce some errors.*

BIOGRAPHY

J. RUBAS received his tertiary education at the University of Stuttgart and the Royal Melbourne Technical College. He joined the Postmaster-General's Department in 1953 and for the first few years worked on telephone exchange equipment installation and maintenance in Melbourne. In 1959 he was transferred to Headquarters and commenced work on traffic engineering problems. Between 1959 and 1962 he was secretary of the Traffic Engineering Committee which prepared and issued the new series of traffic engineering instructions. He continued to work in this field during the subsequent years and investigated a variety of traffic engineering problems. He attended and presented papers at the 4th, 5th, 6th, 7th and 8th International Teletraffic Congresses. These papers, except the first, have been published in the A.T.R.; in addition, he has contributed several articles on traffic engineering to other journals. At present Mr. Rubas is Supervising Engineer of the Telecom Australia Headquarters Traffic Engineering Section.

Computations with Smooth Traffics and the Wormald Chart

D. T. NIGHTINGALE
Telecom Australia, Sydney, Australia

ABSTRACT

The paper presents an evaluation of the mathematical expressions for overflow traffic when the number of circuits assumes real quantities. Efficient computational methods are given for evaluation of overflow traffic to an arbitrary accuracy.

The resulting methods are applied to Wilkinson's equivalent random theory (ERT-W) for both rough and smooth traffics. A new ERT(ERT-N) is derived and applied which avoids the iterative complications of ERT-W. An equivalent binomial theory for rough and smooth traffics is defined and results compared with the ERT's and exact results. An equivalent random queue theory is defined for several queue disciplines and a brief comparison made with exact results. A directly computed Wormald chart and a table of values for E(-1,A) and $\psi(0,A)$ are included where $\psi(0,A) = (\partial \ln \Gamma(N+1,A)/\partial N)_{N=-1} - \ln A$.

INTRODUCTION

In a plane formed with N as abscissa and A as ordinate, rough traffic problems of an Erlang overflow type occur in the first quadrant and smooth traffic problems in the second. This paper is primarily devoted to evaluation of the traffic quantities of interest in the second quadrant. However since analytic continuation exists between the first and second quadrants the general results will be applicable in both quadrants.

The prime quantities of interest are the cumulants of overflow traffic when a pure chance traffic is applied to a group of N circuits all fully available. Wilkinson's equivalent random theory involves the use of the mean and variance only to determine an equivalent group of N circuits and equivalent pure chance traffic A to convert a problem with non-random traffic input to one of simple full availability.

When the traffic is smooth new methods of computing the mean overflow traffic are required, for this condition results in a negative equivalent choice. The iterative complications of the Wilkinson method are still present however so the author has derived a new equivalent random theory involving three cumulants, which enables solutions to be obtained without iteration.

Traffic arising from a limited number of sources exhibits smooth characteristics and theoretical models for a number of practical situations have been evolved in both congestion and queueing theory. These models, on the one hand, enable an equivalent binomial theory to be devised to approximate infinite source models and, on the other, as checks on the application of infinite source models to limited source problems. To this end an equivalent random theory for queues has been studied briefly.

THE ERLANG LOSS FORMULA IN CONTINUOUS FORM

Equation (1) is the usual form of Erlang's Loss formula for full availability systems where B(N,A) is the probability of loss, N, a positive integer, the number of circuits.

$$B(N,A) = \frac{\frac{A^N}{N!}}{\sum_{i=0}^{N} \frac{A^i}{i!}} \qquad \text{..........(1)}$$

The denominator of (1) may be classed as an incomplete exponential function, $f_1(N,A)$, of the first kind since its terms correspond to those of the exponential series. Clearly, when N is large enough, $f_1(N,A) \to \exp(A)$ and leads to the Poisson approximation to B(N,A).

When N is continuous, the factorial of N(N!) is usually represented by the gamma function $\Gamma(N+1) = \gamma(N+1,A) + \Gamma(N+1,A)$, where $\gamma(N+1,A)$ and $\Gamma(N+1,A)$ are incomplete gamma functions of the first and second kinds, defined by the integrals (2) and (3), respectively,

$$\gamma(N+1,A) = \int_0^A \exp(-t).t^N.dt \qquad \text{..........(2)}$$

$$\Gamma(N+1,A) = \int_A^\infty \exp(-t).t^N.dt \qquad \text{..........(3)}$$

Integration of (3), by parts, leads to the result $\Gamma(N+1,A) = f_1(N,A).\,\Gamma(N+1).\,\exp(-A)$ which, on insertion into (1) gives (4), the continuous form of the Erlang loss formula.

$$B(N,A) = \frac{A^N.\exp(-A)}{\Gamma(N+1,A)} \qquad \text{..........(4)}$$

The traffic, m(N,A), overflowing from a full-availability group is obtained from (5) with variance v(N,A) from (6).

$$m(N,A) = \frac{A^{N+1}\exp(-A)}{\Gamma(N+1,A)} \qquad \text{..........(5)}$$

$$v(N,A) = m(N,A).\left[1-m(N,A) + \frac{A}{N+1-A+m(N,A)}\right] \qquad \text{..........(6)}$$

The exponential integral $E_n(Z)$, defined by the integral (7), is related to $\Gamma(N+1,A)$. For, by setting Z = A and Zt = W, (7) reduces to $A^{n-1}.\Gamma(1-n,A)$ which, on substitution in (5), provides the interesting result (8).

$$E_n(Z) = \int_1^\infty \frac{\exp(-Zt)}{t^n}.dt \qquad \text{..........(7)}$$

$$m(-n,A) = \frac{\exp(-A)}{E_n(A)} \qquad \text{..........(8)}$$

The function $E_n(A)$ (not to be confused with E(N,A)) has been extensively tabulated so that results are readily derivable for B(N,A) with negative argument N. Similar functions are available (Ref. 5), which enable B(N,A) (and m(N,A)) to be obtained in the third and fourth quadrants of the A/N plane although they are not known to have any practical traffic significance.

Various regions of the second quadrant of the A/N plane are of interest because of their relationship to well-known functions. For example setting n=1 in (7) one obtains the exponential integral $E_1(A)$ which is variously represented by $\Gamma(0,A)$ or $-Ei(-A)$.

Brettschneider (Ref.2) chose a rational approximation to $-Ei(-A)$ and, in conjunction with the inverse of (9), generated values of m(N,A) for negative integral N.

$$m(N+1,A) = \frac{A.m(N,A)}{N+1+m(N,A)} \qquad \text{..........(9)}$$

A function, $G_p(W)$, represented by the integral (10), is of special interest. Comparison of (10) with (2) shows it to be identical in form to the incomplete gamma function of the first kind. Hence (11) follows readily from (3), (5) and (10).

$$(\tfrac{1}{p})!\,G_p(W) = \int_0^W \exp(-t^p).dt = \frac{1}{p}\int_0^{W^p} \exp(-v).v^{(\frac{1}{p}-1)}\,dv \qquad \text{..........(10)}$$

$$m(\tfrac{1}{p}-1,A) = \frac{A^{\frac{1}{p}}.\exp(-A)}{\Gamma(\frac{1}{p})(1-G_p(p\sqrt{A}))} \qquad \text{..........(11)}$$

When p=2, $G_p(W)$ reduces to the error function erf($\sqrt{A}$), in which case (11) reduces to (12) where erfc($\sqrt{A}$) is the complementary error function (1-erf($\sqrt{A}$)).

$$m(-\tfrac{1}{2},A) = \sqrt{\frac{A}{\pi}}\cdot\frac{\exp(-A)}{\mathrm{erfc}(\sqrt{A})} \qquad \text{..........(12)}$$

Jagerman (Ref. 12) gives asymptotic forms for $B(-\frac{1}{2},A)$ and B(-1,A) mainly derived from Whittaker functions. Extensive work has been done by Akimaru and Nishimura (Refs. 6 and 7) on the differential coefficients of the Erlang function, when N is restricted to a positive real variable. However since analytic continuation exists between the first and second quadrants of the A/N plane, the basic results, (13) and (14), are also valid in the second quadrant.

$$\frac{\partial}{\partial A}.B(N,A) = B(N,A).\left[\frac{N}{A} - 1 + B(N,A)\right] \qquad \text{..........(13)}$$

$$\frac{1}{B(N,A)}.\frac{\partial}{\partial N}.B(N,A) = \ln A - \frac{\partial}{\partial N}\ln\Gamma(N+1,A) = -\psi(N+1,A) \qquad \text{..........(14)}$$

Direct differentiation of (9) results in (15) and, in conjunction with (14), one obtains the recursive function (16) for $\psi(N+1,A)$.

$$D_1 = \frac{B_1}{B_o}.\left[D_o(1-B_1) - \frac{B_1}{A}\right] \qquad \text{..........(15)}$$

$$\psi(N+2,A) = \frac{B_1}{A.B_o}.\left[(N+1).\psi(N+1,A)+1\right], \qquad \text{..........(16)}$$

where $D_i = \frac{\partial}{\partial N}.B(N+i,A)$ and $B_i = B(N+i,A)$.

Inverse formulae for (15) and (16) are readily derivable.

The smallest positive value of N, for which a differential coefficient may be required is zero, which requires $\psi(1,A)$ in (14). Since $\psi(1,A)$ is, conveniently $\exp(A).E_1(A)$ one obtains the result (17) for D_o, from (8) and (14).

$$\left[\frac{\partial}{\partial N}.B(N,A)\right]_{N=0} = -\frac{1}{m(-1,A)} \qquad \text{..........(17)}$$

In applying the inverse of (15) or (16) through zero, with integral values of N, a singularity occurs. Hence for this case $\psi(0,A)$ is needed with generation commencing from N = -1. $\psi(0,A)$ is also needed for evaluation of second differential coefficients (Ref.7).

EQUIVALENT RANDOM THEORIES

The original equivalent random theory (ERT-W) devised by Wilkinson (Ref.3), treats N as a positive real variable in which the mean overflow traffic is given by (5) with variance (6). When values of m and v are specified which result from a simple full availability group, with a single pure chance traffic offered, simultaneous solution of (5) and (6) precisely determine the values of A and N concerned. If v is less than m the traffic is smooth and solution of (8) and (6) will produce an exact negative value of N. When an offering stream arrives from several different sources, each in a different stage of degeneration (from pure chance traffic) the ERT produces a single stream equivalent of the arriving streams irrespective of differences in the higher moments.

From Wallstrom (Ref.15) the factorial moments, Fq, of traffic overflowing a full availability group are defined in (18), in terms of Kosten polynomials R(N,r). Since individual polynomials sum to zero when N is negative and are undefined when N is not integral it is necessary to define a new function Z(N,a,b) as a ratio (19).

$$F_q(N) = A^q\,\frac{R(N,0)}{R(N,q)} \qquad \text{..........(18)}$$

$$Z(N,a,b) = \frac{R(N,a)}{R(N,b)} \qquad \text{..........(19)}$$

Equations (20) to (22) are among a number of recurrence relationships satisfied by R(N,r)(Ref. 15), which are used to derive (23) for Z(N,a,b).

$$R(N,r) = R(N-1,r) + R(N,r-1) \qquad \text{..........(20)}$$

$$N.R(N,r)=(A+N-1+r).R(N-1,r)-A.R(N-2,r) \qquad \text{......(21)}$$

$$N.R(N,r)=A.R(N-1,r)+r.R(N-1,r+1) \qquad \text{..........(22)}$$

$$\frac{1}{Z(N,0,q+1)} = \frac{1}{q}\left[\frac{N+q-A}{Z(N,0,q)} + \frac{A}{Z(N,0,q-1)}\right] \qquad \text{......(23)}$$

Clearly, from (19), Z(N,0,0) = 1 and since $F_1(N)$=m(N,A), Z(N,0,1) = B(N,A). Hence, if $F_1(N)$ is available for any values of A and N, Z(N,0,q) is available for all q, from (23). Z(N,0,q) may be eliminated from (23) to produce (24) in terms of F_q so that, given A, N and m(N,A) all factorial moments of the overflow traffic may be generated, irrespective of the character of N.

$$\frac{A}{F_{q+1}(N)} = \frac{1}{q}.\left[\frac{N+q-A}{F_q(N)} + \frac{1}{F_{q-1}(N)}\right] \qquad \text{..........(24)}$$

A closed solution, for A and N may be obtained from (24), when F_1, F_2, and F_3 are known. Setting ζ = (N-A) and q=2,3 in (24), ζ is given by (25), from which A and N follow directly.

$$\zeta = \frac{F_3(2F_1 + F_2) - 2F_2^2\,(F_1+1)}{2F_2^2 - F_1F_3} \qquad \text{..........(25)}$$

whence,

$$A = \frac{F_2}{F_1}\,(\zeta + 1) + F_2 \qquad \text{..........(26)}$$

$$\text{and } N = \zeta + A \qquad \text{..........(27)}$$

Limiting properties of (25) are of interest particularly if F_1 and F_2 are assumed constant. The full extent of the variation of F_3 produces regions in which both positive and negative values of A and N occur. For the purposes of the proposed ERT(ERT-N) and ERT-W, F_3 lies between the values given by (28) and (29) in which region A remains positive.

$$\zeta = 0 \text{ when } F_3 = \frac{2F_2^2\,(1+F_1)}{2F_1 + F_2} \qquad \text{..........(28)}$$

$$\zeta = +\infty \text{ when } F_3 = \frac{2F_2^2}{F_1} - \varepsilon \qquad \text{..........(29)}$$

In principle any arrival process, for which F_1, F_2 and F_3 are specified, may be approximated by a single Erlang stream. ERT-N gives an exact result when the arrival process consists of the overflow traffic from a single full-availability group offered pure chance traffic. Where the overflow traffic is derived from a number of such sources, for which the originating parameters are known, ERT-N is an approximation to the true result, is directly calculable and, in most cases, is more accurate or safer than ERT-W. Most practical traffic calculations commence at some point, with an assumed pure chance traffic, therefore the iterative intricacies of ERT-W can be avoided by progressively deriving the first three cumulants of the traffic overflowed (or carried) by successive groups of circuits. If the first three cumulants are m(mean), v(variance) and w, the factorial moments are given by (30) to (32).

$$F_1 = m \qquad \text{..........(30)}$$

$$F_2 = v - m + m^2 \qquad \text{..........(31)}$$

$$F_3 = w + 3mv + m^3 - 3F_2 - m \qquad \text{..........(32)}$$

When deriving the factorial moments of several streams the values m, v and w, to be used in (30) to (32) represent the sum of the separate cumulants in the arriving streams. In cases where only m and v are known, for an arriving stream, recourse may be had to ERT-W for an approximation to w.

The resultant overflow traffic from circuits following an equivalent group may be obtained from the recurrence (9). Alternatively it can be shown, from (22), that the moments overflowing successive circuit groups are given by (33).

$$\frac{1}{F_q(N+1)} = \frac{1}{F_q(N)} + \frac{q}{F_{q+1}(N)} \qquad \text{..........(33)}$$

COMPUTATION OF m(N,A)

Practical uses for the Erlang formula seldom occur much outside of the triangle bounded by N = 0 and N = A, when N is positive. The overflow traffic m(N,A) becomes vanishingly small for quite modest extensions of N and a problem does not exist when A = 0. The condition N = 0 is often used in computer programmes to allow traffic to overflow without change of properties. If N is finite and A →0, m(N,A) →0 for all positive values of N. Hence, in the first quadrant, the +N axis represents the limit m(N,0) = 0 and the +A axis m(0,A)= A.

In the second quadrant the line N = -1 is described, in (8), by the exponential integral $E_1(A)$, which may be represented by the series expansion (34).

$$E_1(A) = -(\gamma + \ln A) - \sum_{p=1}^{\infty} \frac{(-A)^p}{p!p} \qquad \text{..........(34)}$$

where γ = Euler's constant = 0.5772156649...

For very small traffics, A, $E_1(A) \to -(\gamma + \ln A)$ and as A→0, $E_1(A)$ rises slowly to infinity; e.g. when A is as little as 10^{-99}, $E_1(A) = 225.076$. Hence, from (8), m(-1,A) →0 as A→0 and, at this point, from (6), v(-1,0) = 0. It may be shown generally that m(-N,0) = N-1 and v(-N,0) = 0, i.e. the negative N axis exhibits a finite mean with zero variance at the integral points. For the region -1<N<0, Gp(0) = 0 and $m(\frac{1}{p} - 1, 0) = 0$ for any real p. Similarly $m(\frac{1}{p} - 1 - f, 0) = f - \frac{1}{p}$, f integral. Elsewhere in the second quadrant the variance is always less than the mean m, and A is less than v. It may be shown, easily, that v→0 as A→0, for any n(= -N). By introduction of the marginal occupancy (Ref. 8) h_N=m(N,A) - m(N+1,A), (6) can be re-expressed in the alternative form (35) and since from above, m(-N,0) = N-1, it follows that v(-N,0) = 0.

$$v(N,A) = m(N,A)^2 \left[\frac{1}{h_N} - 1 \right] \qquad \text{..........(35)}$$

The continued fraction (CF) representation of $E_n(A)$ (n = -N) is a useful form for direct computation of m(-N,A) over most regions in the second quadrant of the A/N plane. The general form of the CF to be used is given in (36) and a computational process in (37).

$$m(-N,A) = a_0 + \frac{b_0}{a_1+} \cdot \frac{b_1}{a_2+} \cdot \frac{b_2}{a_3+} \cdot \text{---} \qquad \text{..........(36)}$$

where, with q an arbitrary integer,

$$a_{2q} = A,\ a_{2q+1} = 1,\ b_{2q} = N+q,\ b_{2q+1} = q+1.$$

Hence, setting $Q_{q+1} = 1$, computation of (36) may be arranged in the sequence (37).

$$Q'_q = \frac{N+q}{Q_{q+1}} + A; \quad Q_q = \frac{q}{Q'_q} + 1;$$

$$Q'_{q-1} = \frac{N+q-1}{Q_q} + A; \quad Q_{q-1} = \frac{q-1}{Q'_{q-1}} + 1;$$

etc., which leads to $m(-N,A) = Q'_0$ ....(37)

It is found that the best result for m(-N,A), using 10-digit, floating-point computation, may be obtained with q approximated by K/√(abs N+A). For the region A≥0.5, K = 115 and for N≤-10, K = 259. Precise accuracy in q is not necessary for a precise result in m provided that the value of q, given by the chosen approximation, is large enough.

In the remaining region of the second quadrant the value of q becomes too large for practical purposes so that other methods must be used. For the region n<1 equation (11) is suitable, with G_p computed from the expansion (38) and $\Gamma(\frac{1}{p})$ obtained from a Sterling process, (39) or other suitable rational approximation.

$$(\tfrac{1}{p})!\,Gp(p\sqrt{A}) = p\sqrt{A}\,.\sum_{j=0}^{\infty} \frac{(-A)^j}{j!(jp+1)} \qquad \text{..........(38)}$$

$$\Gamma(\tfrac{1}{p}) \doteqdot \frac{\Gamma(r)}{\prod_{i=0}^{s-1}(\frac{1}{p}+i)} \qquad \text{..........(39)}$$

where. $r = \frac{1}{p} + s.$

$$\ln\Gamma(r) \doteqdot (r-1/2)\ln r - r + (1/2)\ln(2\pi) + \emptyset(r),$$

$$\emptyset(r) \doteqdot (1 + \frac{1}{z}(1 + \frac{1}{2z}(1 - \frac{1}{15z}(139 + \frac{571}{4z}))))$$

and $z = 12r$.

The parameter s would be chosen at about 30 for the proposed region of application in the A/N plane. The region for A<0.5, n>1 would have m(N,A) obtained by recurrence using the inverse of (9). However some loss of accuracy will occur in recurring across the leading diagonal n(= -N, N+ve)/A. Let n take a small value ∝, then m(∝,A) will lie close to and above A, (A+δ) say. If the inverse of (9) is now applied to find m(∝-1,A) the result is ∝.(A+δ)/δ. Hence the accuracy of recurrence is dependent upon the accuracy with which δ is known. Since δ<<A, and (A+δ) occupies all digits available, accuracy is lost in recurring across the diagonal. For example at A=0.5 and with 10-digit computation, the error in m is 30 in 10^{-10} at ∝=.01 rising to 8670 in 10^{-10} at ∝=.0001.

On the general question of recurrence, Miller (Ref.10) states that accuracy is lost in recurring towards the leading diagonal in the second quadrant and gained in recurring away. Rapp (Ref.9) proved the latter part of this rule, by algebraic means, for increasing recurrence in the first quadrant. The first quadrant has also been examined by Levy-Soussan in terms of a CF solution (Ref. 14). The CF process (37) is also valid in the first quadrant of the A/N plane and is a terminating CF at q=N for positive integral values of N. For real N(+Ve) q varies significantly along the diagonal A/N but may be approximated by q=5(25 - √((A+5)/(N+5))) in the range A≥N≥1. When A<N, m(N,A) would be obtained from within the region bounded by A>N and recurrence used to the point in question. Direct computation is indicated for the region N<A<1 using an increasing series for the incomplete gamma function of the first kind (e.g. (40)) together with a Sterling approximation for Γ(N) in (5).

$$\gamma\,(N,A) = \frac{A^N}{N}\left[1 + \sum_{r=1}^{\infty} \frac{(-A)^r.N}{r!(N+r-1)}\right] \qquad \text{..........(40)}$$

Direct evaluation of m(N,A), N real, enables differential coefficients to be readily obtained. ∂B(N,A)/∂A is available directly from (13), whereas ∂B(N,A)/∂N requires a value for $\Psi(N+1,A)$, which is computationally difficult (Refs. (6) and (7)). Using a 6-point Lagrange formula one may derive approximation (41) which exhibits an error of the order of $h^6\Delta^6/30$.

$$\left[\frac{\partial}{\partial N}m(-N,A)\right]_{N=S} \doteq \frac{1}{60h}\sum_{r=0}^{5} C_r.m((1-r).h-S,A) \quad \ldots\ldots\ldots.(41)$$

where $C_0 = +12$, $C_1 = +65$, $C_2 = -120$,

$C_3 = +60$, $C_4 = -20$, $C_5 = +3$.

TABLE 1

A	B(-1,A)	PSI(0,A)
.05	7.70882132	1.49819312
.10	4.96365970	1.21351667
.15	3.91819963	1.05587744
.20	3.34817973	.94881104
.25	2.98310345	.86883012
.30	2.72657362	.80564029
.35	2.53507355	.75382743
.40	2.38588774	.71020120
.45	2.26592114	.67272827
.50	2.16705706	.64003589
.55	2.08397797	.61115538
.60	2.01304428	.58537879
.65	1.95167489	.56217381
.70	1.89798478	.54113058
.75	1.85056287	.52192718
.80	1.80833030	.50430642
.85	1.77044712	.48805977
.90	1.73624885	.47301597
.95	1.70520247	.45903275
1.00	1.67687503	.44599071
1.50	1.48724316	.35072295
2.00	1.38378190	.29171139
2.50	1.31784498	.25086193
3.00	1.27185812	.22063233
3.50	1.23782650	.19723059
4.00	1.21155934	.17851278
4.50	1.19063615	.16316409
5.00	1.17355619	.15032835
6.00	1.14730768	.13003073
7.00	1.12804727	.11466703
8.00	1.11329178	.10260971
9.00	1.10161567	.09288275
10.00	1.09214022	.08486310
12.00	1.07768655	.07240290
14.00	1.06717293	.06315913
16.00	1.05917656	.05602238
18.00	1.05288760	.05034278
20.00	1.04781065	.04571370

The first order approximation (42) may suffice for many practical applications and, with h=1, is the form used by Rapp (Ref.9) for iterative evaluation of ERT-W.

$$\frac{\partial}{\partial N}m(N,A) \doteq \frac{1}{2h}\left[m(N+h,A) - m(N-h,A)\right] \quad \ldots\ldots\ldots.(42)$$

Equation (16) may be used for negative recursive evaluation of the differential coefficient except when passing through N=0 at the integral point. An increasing series (43), has been used to compute Table 1 giving values of $\psi(0,A)$ up to A=20 which necessitates the use of 29-digit arithmetic to obtain 9-digit accuracy. To compute the values of $E_1(A)$ to the required accuracy a sequence due to Miller and Hurst (Ref.11) was used.

$$\psi(0,A) = \frac{1}{E_1(A)}\left[\frac{1}{2}(\gamma + \ln A)^2 + \frac{\pi^2}{12} + \sum_{k=1}^{\infty}\frac{(-A)^k}{k!k}\right] \ldots(43)$$

For values of A>20 (approx) the asymptotic form (44) may be used in which, with terms continued to $p = [A]$, the error in $\psi(0,A)$ approximates $p!/A^{p+3}$.

$$\psi(0,A) \sim A.m(-1,A).\sum_{s=2}^{p+1}\left(\frac{-1}{A}\right)^s\frac{(s-2)!}{m(-s,A)} \quad \ldots\ldots\ldots.(44)$$

NUMERICAL CONSIDERATIONS

In 1968 Wormald published (Ref.4) the results of a simulation exercise in which congestion resulting from offered smooth traffics were obtained. The results are principally concerned with the second quadrant of the A/N plane and represent a continuation of the well known Wilkinson chart (Ref.3). The Wormald chart combines both the first and second quadrants, of the A/N plane, in one quadrant in such a way that the leading diagonal' represents pure chance traffic. A directly computed version of this chart is given in Fig.1 which was derived from the processes described in this paper. To read the chart one notes that the radials represent pure chance traffics and circuits count negatively from the diagonal upwards, positively downwards. Comparison with the originally simulated version shows remarkably close agreement over the whole range. Ample verification is thus provided of the theoretical processes which describe the degeneration, from pure chance, of one infinite source traffic stream overflowed from a real quantity of circuits, all fully available.

Practical application of ERT-W (and ERT-N when F_3 is not known) requires the simultaneous solution of (5) and (6) to determine equivalent values of A and N given the offered resultant traffic m and v. The author uses (45) (a modified form of the Rapp approximation (Ref. 9)) to obtain a first approximation A* to A; N* is then obtained from (46) which is an inversion of (6). The problem is thus essentially of one variable, in A.

$$A^* \doteq v + (1+Q+Q^2)\frac{v}{m}\left(\frac{v}{m} - 1\right) \quad \ldots\ldots\ldots.(45)$$

where $Q = 1 - 1/(m + v/m)$

$$N^* \doteq A^*/Q - m - 1 \quad \ldots\ldots\ldots.(46)$$

Various methods are in use for refining the value of A*, for example, the Newton-Raphson iteration described by Ott and De Los Rios (Ref.13) and the Reguli Falsii method described by Rapp (Ref.9). The author uses the former for computer evaluation and the latter for manual computation. Equations (45) and (46) are applicable over both quadrants and provide better accuracy in the second quadrant than the Rapp equation and acceptable accuracy elsewhere. Some examples follow:-

N	A	A*Rapp	A*Equation (45)
-.3	.05	-.0082	.0615
-.3	.50	.4358	.5224
-.3	2.50	2.4728	2.5054
-.3	4.95	4.9361	4.9510
-5.0	5.00	4.9325	5.0016

Traffic originating from a limited number of sources S, each offering 'a' erlang/source (average), will have a mean value of m = S. a erlang and a variance of v=Sa(1-a). In the limit, as $S\to\infty$, the traffic changes from smooth (v<m) to pure chance (v=m). Hence it may be expected that models based on a limited number of sources would have a congestion response which is different to that predicted by either ERT-W or ERT-N. Hence application of an ERT to a limited source situation, will produce a different congestion to a limited source model and the difference may be expected to diminish as S becomes large. The Engset model, for call probability of loss, is given in (47) which may be economically computed by the sequence (48).

$$E(N,b) = \frac{\binom{S-1}{N}b^N}{\sum_{j=0}^{N}\binom{S-1}{j}b^j} \quad \ldots\ldots\ldots.(47)$$

where b = a/(1-a(1-E(N,b))).

$$E(N,b) = 1/\Sigma T_R, \quad \ldots\ldots\ldots.(48)$$

where

$T_0 = 1$, R = 1 to N, E = 0 initially and

$$T_R = T_{R-1}\frac{(N-R+1)}{(S-N-1+R)}\cdot\frac{1}{b}.$$

Sequence (48) is used iteratively, by adjusting b, until the change in E is as small as desired. The same sequence may also be used to compute time congestion by replacing (S-1) by (S), which, when further multiplied by (S-N)/N, produces traffic congestion. It will be noted that the first iteration of (48) produces the equivalent O'Dell result.

Mina (Ref.1) proposed an equivalent binomial theory (EBT) based on the binomial formulae for both smooth and rough traffics. Since the mean and variance of a traffic stream are defined in terms of S and 'a', the EBT involves finding ficticious values from the m and v supplied. The binomial models are continuous through S = ∞ so that (48) may be used for both rough and smooth traffics where, for the former case, both S and 'a' are negative. This fact was noted by Wallstrom (Ref. 15, p344) in passing and its significance did not apparently occur to Mina.

A simple trunking system, comprising 8 primary groups of 2 circuits, each offered 2 erlang of pure chance traffic, directs traffic to a secondary group of X circuits. The probability of loss experienced by the secondary group, as predicted by various models, is listed in TABLE 2 when the primary carried traffic is offered to X and in TABLE 3 when the primary overflow traffic is offered to X.

TABLE 2, CARRIED TRAFFIC

MODEL	X =2	5	9	15	EQUATIONS
ERT-N	.8090	.5377	.2335	.0186	(26), (27)
A*/N*	.8089	.5371	.2322	.0180	(45), (46)
EBT	.8118	.5431	.2290	.0023	(48)
SIMULATION	.8161	.5345	.2236	.0014	-

TABLE 3, OVERFLOW TRAFFIC

MODEL	X =2	5	9	15	EQUATIONS
ERT-N	.7459	.4159	.1243	.0050	(26), (27)
A*/N*	.7458	.4152	.1233	.0048	(45), (46)
EBT	.7365	.3959	.1115	.0056	(48)
SIMULATION	.7407	.4005	.1165	.0039	-

Kibble (Ref.20) provides exact and ERT-W solutions for a large number of gradings. In every case ERT-N is higher than ERT-W and where ERT-W is below the exact solution ERT-N is, in some cases, closer to the exact solution and in others slightly higher. The elementary two-group grading with three outlets has been solved exactly and some sample comparisons with ERT-N are given below, with a balanced offered traffic of A erlang.

A	E(exact)	E(ERT-N)
.1	.00130	.00130
.5	.03153	.03178
1	.10494	.10606

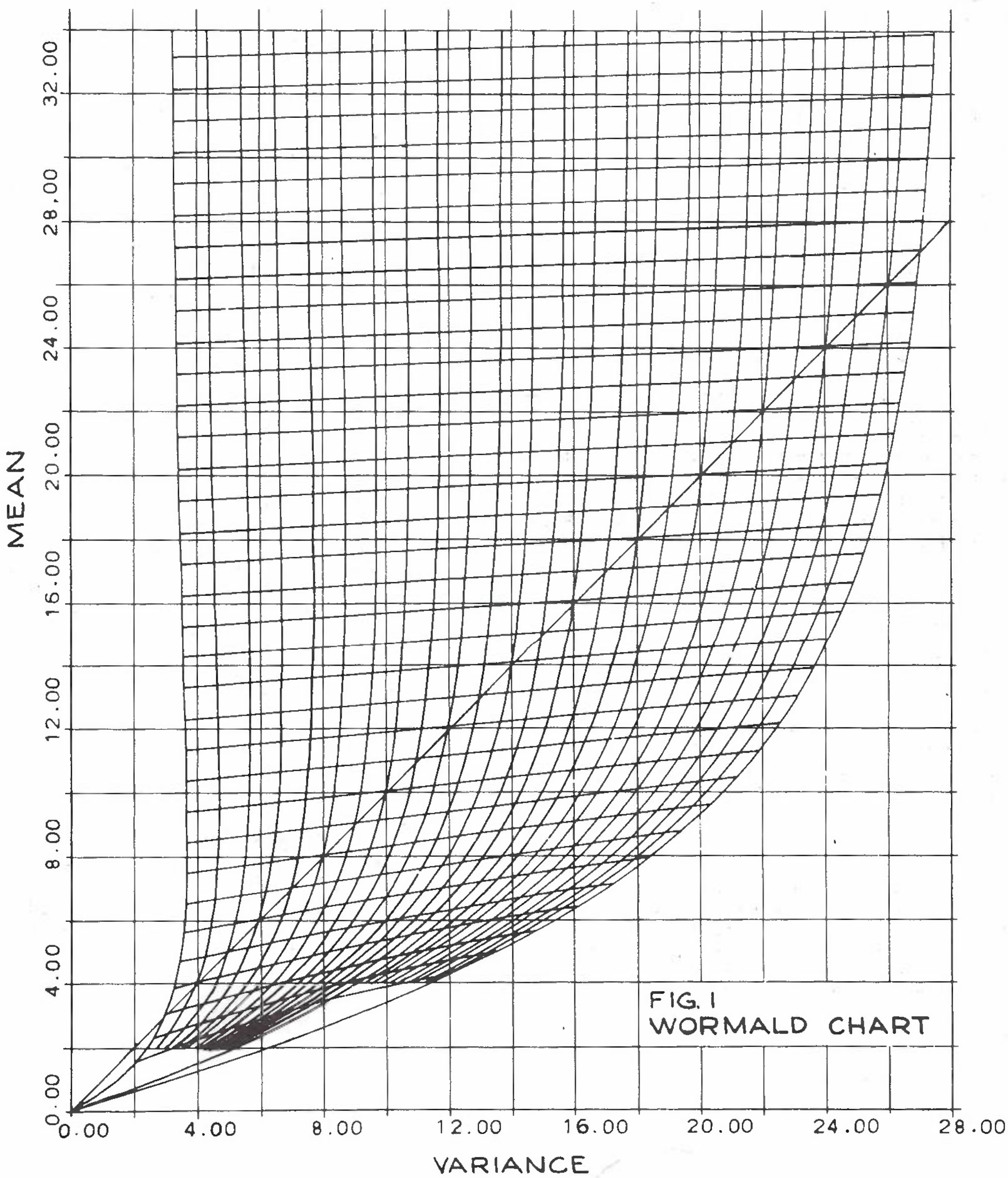


FIG. 1
WORMALD CHART

2	.26705	.26923
3	.39600	.39810
5	.56199	.56328

Schehrer (Ref. 16) has provided an exact solution for consecutive Engset groups an example for which has S=40, b=0.5 offering to a primary group of 10 circuits followed by a secondary group of 10 circuits. The ERT solution (with N negative) gives 4.4294 erlang overflowed to the secondary group (4.6249 exact) and 0.1273 lost (.1201 exact). The EBT gives 4.6674 and .120 respectively when the offered traffic 'a' is assumed constant for the intermediate group calculation.

One program tape for an HP65 calculator can accommodate equations (37), (6) and (9) plus its inverse and a typical listing is given in Table 4. With this program one may evaluate, to nearly 10 digit accuracy, m(N,A) and v(N,A) over the practical range of interest in A and real N. All other equations given in this paper have been directly applied to the HP65 calculator with the exception of (43) which requires double precision arithmetic. It is noted that Brettschneider (Ref.21) suggests integer N program forms for the HP65 for (1), (6), (9), (47) and (49).

TABLE 4

HP65 Routine For Equations (37), (6) and (9)

For m(N,A,q): A↑N↑q key A. Any time after an m(N,A) has been obtained:- m(N+R,A): R STO8 key D; m(N-R,A): R STO8 key C; v(N,A): key B. Stores used are: 1A, 2N, 4m, 6v and 8.

f	RCL8	g	B	1	1	GTO	÷
PROG	↑	DSZ	RCL1	+	+	C	RCL1
LBL	↑	GTO	↑	RCL4	↑	RTN	X
A	RCL2	1	RCL4	X	RCL4	LBL	STO4
STO8	–	RCL2	1	STO6	X	D	g
g↓	RCL4	CHS	+	RTN	RCL1	RCL4	DSZ
STO2	÷	RCL4	RCL2	LBL	↑	↑	GTO
g↓	RCL1	÷	+	C	RCL4	↑	D
STO1	+	RCL1	RCL1	RCL2	–	RCL2	RTN
1	÷	+	–	↑	÷	1	gNOP
STO4	1	STO4	÷	1	STO4	+	gNOP
LBL	+	RTN	RCL4	–	g	STO2	–
1	STO4	LBL	–	STO2	DSZ	+	–

EQUIVALENT RANDOM QUEUES

Erlang's second formula, for delay working, is usually expressed in the form (49) where N (an integer) and A have the usual meaning. The quantity $B_2(N,A)$ is the probability of a delay occurring and $m_2(N,A)=A.B_2(N,A)$, the delayed traffic.

$$B_2(N,A)=\frac{\frac{A^N}{N!}\cdot\frac{N}{N-A}}{\sum_{p=0}^{N-1}\frac{A^p}{p!}+\frac{A^N}{N!}\cdot\frac{N}{N-A}} \qquad \text{..........(49)}$$

One may develop an equivalent random theory for queues (ERTQ) by converting (49) to its equivalent form (50) with real N, which is achieved by replacing the sum term in (49) by its equivalent from (1).

$$m_2(N,A) = \frac{N.m(N,A)}{N-A+m(N,A)} \qquad \text{..........(50)}$$

In the form (50) m_2 may be evaluated for N a continuous real variable so that a non-random source may be approximated by an equivalent pure chance traffic and an equivalent choice. The delay performance for X(say) circuits which succeed the equivalent choice is then obtained simply from $B_2(X,A) = m_2(X+N,A)/m(N,A)$.

The limited source equivalent of (49) may be expressed in the form (51) (Ref. 18) which can be computed by the sequences (52) and (53).

$$B_2(X,S,a) = \frac{T}{T+R} \qquad \text{..........(51)}$$

where $T = \sum_{q=0}^{S-X} T_q, T_q = \binom{S}{X+q}\frac{(X+q)!}{X!X^q}.b^{X+q}$

and $R = \sum_{p=0}^{X-1} R_p,\ R_p = \binom{S}{p}b^q.$

$$R_o = 1,\ R_p = R_{p-1}.\frac{S-p+1}{p}.b \qquad \text{..........(52)}$$

$$T_o = R(p=X),\ T_q = T_{q-1}.\frac{S-X-q+1}{X}.b \qquad \text{..........(53)}$$

The quantity b assumes different values to those applying to the Engset model because the State of free sources is different. If m(S,a) is the mean of the offered traffic which is known, b is obtained, by iteration, from (54).

$$m(S,A) = S.a = \sum_{p=o}^{N-1} p.R_p + N.T \qquad \text{..........(54)}$$

Table 2 contains a few-selections of traffics with different variance to mean ratios in which the probability of delay, $B_2(X,A)$ from (49), is compared with exact results, $B_2(X,S,a)$ from (51), It remains to be determined whether smooth traffic offered to a queue is better represented by ERTQ or (51) when the traffic originates from an original pure chance traffic.

Table 2

V/M	S	X	$B_2(X,A)$ (49)	$B_2(X,S,a)$ (51)
.52	5	3	.30392	.35373
		4	.08565	.05450
.95	20	4	.01558	.01393
		5	.00264	.00207
.9	20	4	.13667	.15263
		5	.04232	.03815
		8	.00054	.00028
.8	20	5	.42349	.44815
		8	.03094	.02487
		10	.00322	.00162
.7	30	10	.46135	.53624
		12	.14918	.15286
		15	.01723	.01228
		18	.00113	.00036

The voluntary departure queue model is a modified form of the Erlang queue which requires no restrictions on the value of A. The probability of delay $B_3(N,A)$, which includes (49) as a special case, can be expressed in the form (55) (Ref. 19) with N a real variable.

$$B_3(N,A) = \frac{(1+Q).m(N,A)}{A+Q.m(N,A)} \qquad \text{..........(55)}$$

where $Q = \sum_{q=1}^{\infty}\frac{A^q}{K_q},\ K_q = \prod_{r=1}^{q}(N+r.\frac{h}{d}),$

h = mean holding time

d = mean time of departure, if not served.

Equation (55) is in the form in which ERTQ may be applied, for non-random offered traffics, but space does not permit further elaboration.

ACKNOWLEDGEMENT

This paper is published with the permission of the Australian Telecommunications Commission.

BIBLIOGRAPHY

1. R. Mina. Some Practical Applications of Teletraffic Theory. Appendix 1 by Dr. R. Syski. 5th I.T.C., 14-20 June, 1967.
2. G. Brettschneider. Extension of the Equivalent Random Method to Smooth Traffics. 7th I.T.C., p.411, 1972.
3. R.I. Wilkinson. Theories of Toll Traffic Engineering in the U.S.A.. B.S.T.J., Vol.35, p.421, March 1956.
4. E.G. Wormald. Consecutive Overflow Telephone Traffic Graphs. I.E. Aust., E.E. Trans. pp. 168 - 172, Sept. 1968.
5. M. Abromovitz and I. Segun. Handbook of Mathematical Functions. Dover, 1965.
6. H. Akimaru and T. Nishimura. The First Derivatives of Erlang's B Formula. Elect. Comm. Lab. Tech. J., NTT, Extra Issue No.8, 1966.
7. H. Akimaru and T. Nishimura. The Second Derivatives of Erlangs B. Formula. Elect. Comm. Lab. Tech. J. NTT, Extra Issue No.11, 1963.
8. C. Pratt. The Concept of Marginal Overflow in Alternate Routing. Aust. Telecom. Research J., Vol. 1, Nos. 1 and 2, 1967, p.76.
9. Y. Rapp. Planning of Junction Networks in a Multi-Exchange Area 1. General Principles, Ericsson Technics, No.1, pp77 - 130, 1964.
10. G.F. Miller. Tables of Generalised Exponential Integrals. Mathematical Tables, Vol. 3, D.S.I.R., London, H.M.S.O., 1960.
11. J.Miller & R.R. Hurst. Simplified Calculation of the Exponential Integral. Mathematical Tables Aids Computations (now Mathematics of Computation), Vol. 12, pp 187 - 193, 1958.
12. D.L. Jagerman. Some Properties of the Erlang Loss Function. B.S.T.J., March 1975, pp 525-551.
13. K. Ott & F. De Los Rios. Computation of Urban Trunking Networks with Alternate Routing by Computer. Elect. Comm., Vol. 43, No.2, pp 157-162, 1968.
14. G. Levy-Soussan. Calculation of the Erlang Function through a Continued Fraction Algorithm. Elect. Comm., Vol. 43. No.2, pp 165-170, 1965.
15. B. Wallstrom. Congestion Studies in Telephone Systems with Overflow Facilities. Ericsson Technics, No.3, pp 187-351, 1966.
16. R. Schehrer. On the Exact Calculation of Overflow Systems. 6th I.T.C., Sept. 9-15 1970, pp 147/1 - 147/8.
17. B. Marrows. Circuit Provision for Small Quantities of Traffic. Telecom Jnl. of Aust., Vol. 11, No.6, pp 208-211.
18. A. Descloux. Delay Tables for Finite and Infinite-source systems. McGraw-Hill Book Co U.S.A., 1962.
19. A. Elldin & G. Lind. Elementary Telephone Traffic Theory. L.M. Ericsson, 1974.
20. W.F. Kibble. The determination of telephone traffic blocking in groups of trunk circuits. GEC-AEI Telecommunications, No.37, P.12.
21. G. Brettschneider. Die. Verwendbarkeit programierbarer Taschenrechner fur die verkehrsplanung von Nachrichtenvermittlungssystemen NTZ Vol. 28, H.11, pps 386-388, 1975.

Discussion

P. KUHN, Germany : *In the ERT-N method presented in your paper, the group size and the offered traffic of the equivalent primary group are calculated from 3 given moments. This corresponds to a set of 3 equations for 2 unknowns, which obviously in many cases will not have an exact solution. Could the author please make a comment on the accuracy of the first 3 moments of the overflow traffic generated by the equivalent group as compared with the moments of the actual overflow traffic.*

D. NIGHTINGALE, Australia : *Where the factorial moments F_1, F_2 and F_3 are the moments overflowed from a single Erlang group of circuits, offered pure chance traffic, equations (25) to (27) precisely determine the values of A and N. Subsequent degeneration of this traffic will then be determined exactly by the recurrence (9) and condition (5) will be satisfied. Where the offered traffic is other than pure chance in origin ERT-N ignores condition (5) in selecting an approximate Erlang system whereas ERT-W ignores condition (32). Extensive testing against exact solutions and simulated systems reveals that ERT-N is generally better, or safer, than ERT-W, is simpler to use, and, without iterative complication, covers the entire real range of N. Further comparative examples in the table below illustrate the accuracy of ERT-N, where A is the balanced offered traffic to each group of a 2-group grading. E_{IC} is the g.o.s., I being the number of individuals and C the number of commons.*

A	E_{31}		E_{32}		E_{33}	
	EXACT	ERT-N	EXACT	ERT-N	EXACT	ERT-N
1	.0174	.0175	.0040	.0040	.0008	.0008
2	.1155	.1168	.0563	.0576	.0244	.0250
3	.2457	.2475	.1640	.1631	.1021	.1044
4	.3594	.3610	.2770	.2797	.2053	.2083
5	.4492	.4504	.3734	.3755	.3031	.3058

G. KAMPE, Germany : *The diagram on page 5 of your paper, representing a so-called Wormald Chart, contains various curves but not the corresponding parameters. Could you please give an explanation to these curves and their parameters.*

D. NIGHTINGALE, Australia : *The Wormald chart of Fig. 1 has a diagonal line representing pure chance traffic the magnitude of which is given by the equal ordinates. The radials emanating from the origin represent pure chance traffic lines with values determined by the points at which they cross the diagonal. The transerve or nearly horizontal, lines represent unit circuit increments of value zero at the diagonal. Proceeding from this diagonal one counts negatively in an upward direction and positively downwards. In his original presentation it was Wormald's intention to proceed consecutively from any point on the chart, radially downwards, along the appropriate traffic line, by the required number of circuits to be traversed by the offered traffic. For example the point m=19.5, ν=10.7 coincides with the intersection A=10, N=-10 point. Traversing five circuits downwards on the A=10 line gives m=14.7 and ν=10.5.*

BIOGRAPHY

D. NIGHTINGALE began his engineering career in a chemical manufacturing plant in 1936, followed by six years with the Fleet Air Arm and Royal Air Force during the war. He returned to London University after the war and, after graduation in 1950, became a lecturer in engineering and mathematics until joining the Postmaster-General's Department in 1955. He continued lecturing on a part-time basis with the University of New South Wales between 1956 and 1966. His 21 years with the Department were spent wholly with the New South Wales Administration approximately divided into 5 years construction, 11 years planning and 5 years design.

Mr. Nightingale's qualifications include B.Sc.(Eng.), Hons., from London University, M.E. from New South Wales University, Diploma in Engineering from the West Ham College of Technology, London and a Full Technological Certificate in Electrical Engineering Practice from the City and Guilds College, London. He is a Chartered Engineer, Fellow of the Institution of Electrical Engineers, London and Member of the Statistical Society of Australia.

Modular Engineering of Junction Groups in Metropolitan Telephone Networks

J. P. FARR
Telecom Australia, Perth, Australia

ABSTRACT

If a telephone network is designed so that each junction route is dimensioned to the nearest circuit, then when the network is re-dimensioned some time later to cater for changes in point-to-point traffics the new design would probably show that most of the routes should be changed by at least one circuit. On the other hand, if the network is designed so that each route is dimensioned to the nearest preferred modular size (eg 5, 10, 15, 20, etc), then when the network is re-dimensioned some time later for changed traffics it is likely that only a small proportion of routes would need to be changed. This paper gives a comparison of circuit requirements and costs for a real metropolitan network configured with different choices of module size. The paper also includes comparative statistics for different choices of module size on the number of routes which would require a change in size at the time the network goes through successive stages of re-dimensioning to cater for the change in traffic with time. A strong case is made for the adoption of modular engineering as a design principle for metropolitan networks employing alternative routing.

1. INTRODUCTION

The planning engineer entrusted with the task of dimensioning the junction groups which interconnect a network of telephone exchanges seeks to achieve a design which will minimise the total cost of the switching equipment, transmission facilities and external plant whilst also satisfying the Administration's technical criteria for congestion loss, transmission loss, etc. Since present-day metropolitan telephone networks are large, and utilise fairly complicated alternative routing patterns, network design is a complex and lengthy process. Nevertheless, an optimum design for a complete network can be obtained by following a systematic procedure, and the process lends itself to computer-assisted solution (Ref.1).

However, since the number, location and calling habits of telephone users change with time in response to demographic and social factors, the traffic wanting to flow from one exchange to another is not static. It follows that if the network is to continue to meet prescribed standards of congestion loss the quantity of junctions on the interconnecting routes must change. New routes will sometimes be required, existing routes may increase or decrease in size, and some routes should disappear when they are no longer economically justified. Thus the planning engineer's problem is not merely to arrive at a design which is a minimum cost solution at one point in time, but to achieve a framework in which network costs will be minimised over a period of time.

In a network with alternative routing, an increase in point-to-point offered traffic between two terminal exchanges can be catered for in a number of ways, ranging from carrying all of the traffic increment on the direct high usage route linking the two terminals (assuming that there is such a route), through sharing the increase among the route alternatives, to carrying all of the increment on the final choice route. A re-design based on the higher offered traffic would probably indicate some increase in traffic to be carried on the direct, alternative and final choice routes with consequential increases in circuits on all affected routes. More careful analysis could show that the most economical solution after taking once-only installation and administrative costs into account is to carry all the increased traffic on (say) the direct high-usage route, resulting in a change to this route only. Thus, it is not sufficient to produce a network design which, although it will carry the traffic, takes little notice of the existing network of interconnecting routes.

This paper describes a procedure called 'modular engineering' for designing the interconnecting junction routes for a metropolitan telephone network. It will be shown that the modular engineering process can result in a network which costs very little more at the chosen starting date than would a design using traditional methods, and which costs less to update in response to changes in traffic than does a traditional design.

DEFINITION:
Modular Engineering of trunk and junction groups requires that the quantity of circuits initially provided on a new route should be a preferred size and that additions to existing routes should only be made in multiples of a preferred size.

The principles of modular engineering were first described by Levine and Wernander in a 1967 paper which investigated the possible application of the method to some inter-city trunk networks in the USA (Ref.2). The American Telephone and Telegraph Co. has since adopted the method for dimensioning inter-city trunk routes (Ref. 5).

2. OPTIONS AVAILABLE WITH MODULAR ENGINEERING

If modular engineering is to be adopted as a network design principle, policy decisions must be made with regard to the following factors:

2.1 CHOICE OF MINIMUM PERMITTED ROUTE SIZE

For various practical reasons it is usual to set a lower limit (eg 4 circuits) below which a high usage route will not be established. If the traffic does not warrant this minimum number of circuits the traffic is first offered to the alternative route. When routes are dimensioned in increments of one circuit the minimum size of high usage route to be permitted can be set at any desired value. Modular engineering introduces a complication because the minimum size permitted under the module rule may be smaller or larger than the desired minimum route size. Examples of incompatibility are:-

(i) a module size of 6 and preferred minimum route size of 4.

(ii) a module size of 4 and preferred minimum route size of 6.

Such conflicts can be resolved by making the minimum route size equal to the module size, or by using the exact circuit quantity for cases in which the circuit quantity lies in the range between the preferred minimum route size and the next higher modular quantity.

2.2 CHOICE OF MODULE SIZES

Levine and Wernander concluded that for an inter-city trunk network a module which is a multiple or sub-multiple of the group channel size is most appropriate - eg modules of 6, 12 or 24 circuits (Ref. 2).

In metropolitan networks carrier and PCM systems are generally not yet used extensively and a decision on module size is more appropriately influenced by considerations which relate to the switching and terminal equipment rather than to the transmission equipment. The other major consideration to take into account is the relative economics of each possible module size from both present and future view points; as the costs will be influenced by network structure and traffic levels, a special investigation is warranted for each network.

Later in this paper the relative economics of modules of 2, 3, 4, 5, 6, 8 and 10 circuits are compared for one Australian metropolitan telephone network.

There is also a case for using two or more module sizes, thus permitting a module to be chosen which is a function of the size of the route. For example, modules of two could be used on routes up to 10 circuits, modules of four on routes greater than 10 circuits, and modules of six on routes greater than 50 circuits. Such a policy would tend to make the elapsed time between changes to routes more uniform than would be the case with a fixed module size.

2.3 CHOICE OF BREAKPOINT

Having chosen the module size or sizes to be used there is freedom to choose the breakpoint in each range at which the transition will be made from the lower preferred route size to the upper preferred route size. A strong determinant is that a high usage route should not be increased in size until it is fairly certain that the route would still warrant the larger size even if the actual traffic offered to the route falls marginally short of the forecast traffic. Hence the breakpoint should tend towards the upper end of the module range. Furthermore, Levine and Wernander have shown that the optimum value of the breakpoint is about 50-70 percent of the module size and is not critical in this region (Ref 2).

As an example, if the module size is 5, the minimum route size is 3, and the breakpoint is set at 60 percent of the module size, the following table would apply:

Calculated Circuit Quantity (X)	Modular Engineered Value
$0 \leqslant X < 3$	0
$3 \leqslant X < 8$	5
$8 \leqslant X < 13$	10
$13 \leqslant X < 18$	15
etc.	etc.

2.4 POLICY ON FINAL CHOICE ROUTES

There are two ways of treating final choice routes - either provide the exact circuit quantity as calculated, or provide the next LARGER preferred size. There is no reason why the exact quantity should not be provided, but choosing the next larger preferred size has the advantage of

- giving an in-built factor of safety against under estimation of traffics,
- requiring less frequent changes to the size of the final routes, and
- preserving the principle of modular engineering

On the other hand, the grade of service will be better than intended and the number of final route circuits will increase. For any one route the maximum overprovision would equal the module size, but on average the increase will be about 50 percent of the module size. Thus for the network as a whole the circuit penalty through rounding final routes is approximately-

(0.5)(module size)(number of final choice routes)

3. CALCULATING OVERFLOW TRAFFIC FROM MODULAR GROUPS

The circuit quantity to be provided on a high usage route under the modular system will generally differ from the exact quantity calculated from the design data and the overflow to the alternative choice route has to be based on the rounded quantity rather than the exact quantity. Fortunately, the standard dimensioning formulae, graphs or traffic tables can still be used to calculate the moments (mean and variance) of the overflow traffic.

[If the Marginal Occupancy (also called H-factor) method of computing the ideal quantity of circuits on high usage routes is being used, and the mean overflow (α_x) from the ideal circuit quantity (X circuits) has already been calculated, then a close approximation to the mean overflow (α_y) from the rounded quantity of circuits (Y) is given by -

$$\alpha_y = \alpha_x - H(Y-X), \text{ provided } |Y-X| \text{ is small}$$

H is the design marginal occupancy of the route - ie the incremental traffic to be carried on the last circuit when the offered traffic is held constant ($0 \leqslant H \leqslant 1.0$). Since the actual marginal occupancy will vary as the number of circuits varies (see Section 8), this expression is exactly true only if $|Y-X| \leqslant 1$, but is near enough for most purposes if $|Y-X| \leqslant 3$. Hence, if the module size is $\leqslant 5$ and the breakpoint is $\leqslant 60$ percent of the module size, this approximation will cover the cases to be encountered.]

4. BACKGROUND INFORMATION ABOUT THE MODEL NETWORK

Some findings relating to modular engineering will now be given for a particular metropolitan telephone network - Perth, Australia. All input data used in the studies is quite realistic for this network.

Perth has a population of about 800,000 people and there are about 300,000 telephone stations. The telephone network comprises about 60 terminal exchanges, 10 local tandem exchanges and one trunk exchange. Over 80 percent of traffic originates from crossbar exchanges with alternative routing capabilities; the remaining exchanges switch their traffic via routes with a fixed grade of service. There are at present about 900 high usage routes and 200 final choice routes in the network.

The method used by Telecom Australia for dimensioning the circuit requirements from crossbar exchanges with internal link congestion is based on the 'geometric group' model (Ref 3,4). Final choice junction routes are dimensioned for a grade-of-service of .005. In the studies which follow it is assumed that the overall Variance-to-Mean ratio of aggregated traffics on alternative and final routes is 2.0 - this figure being based on measurements of the Variance-to-Mean ratio on existing overflow routes in the Perth network.

Typical alternative routing patterns permitted in the model network are shown in Fig. 1. Note that all routes carry traffic in one direction only and the differences in routing from origins I1, I2 and I3 to destinations J1, J2 and J3 are intentional.

The results discussed below were obtained by dimensioning and costing the network using a computer-based processing system which is regularly used for designing this particular network. The system is described in Ref. 1.

It should be noted that the study results relate only to traffic which is originated or terminated by subscribers in the Perth metropolitan and outer metropolitan area; transit traffic is not included.

5. EFFECTS OF VARYING THE MODULE SIZE

In order to study the effect of varying the module size on network structure and cost the model network was dimensioned and costed a number of times using different modules of practical interest (viz 2, 3, 4, 5, 6, 8 and 10). For comparison, a reference network was engineered

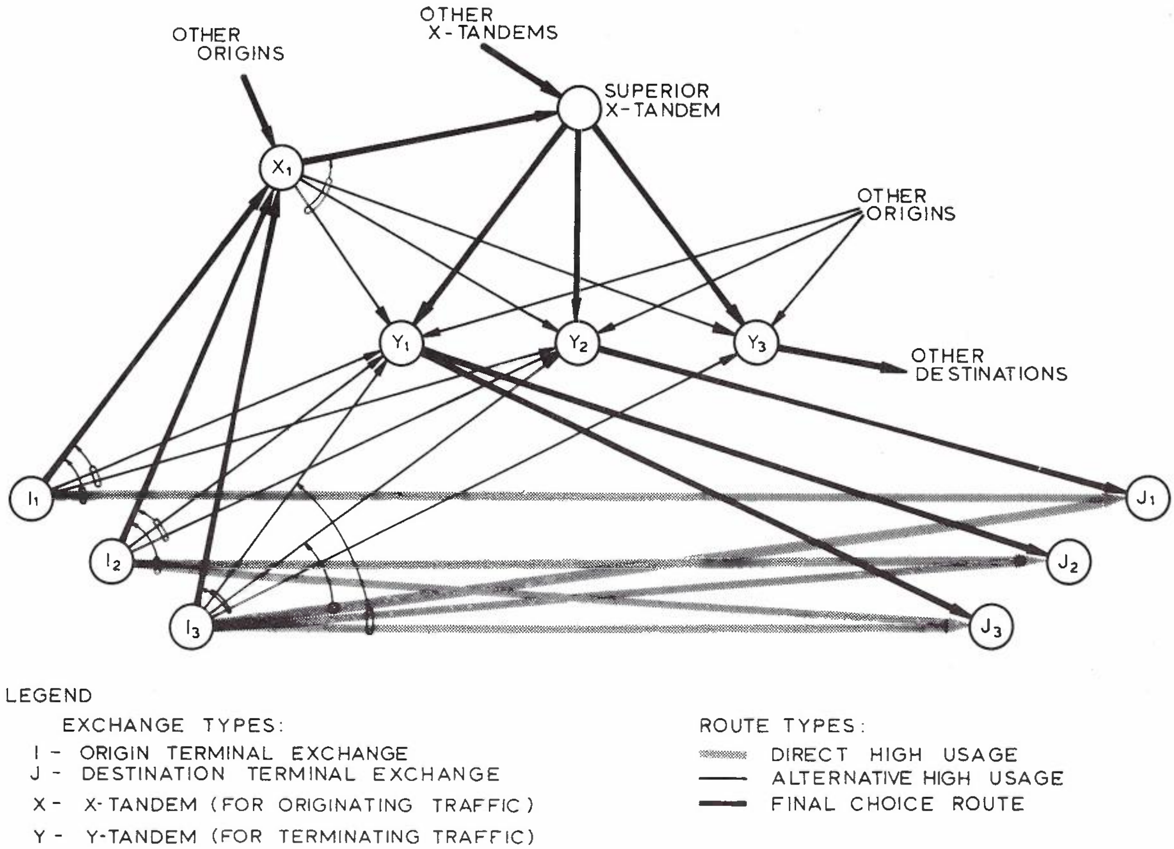


Fig.1 Alternative routing patterns in the model network

by dimensioning each route to the nearest single circuit using standard procedures, as outlined for example in Ref 4.

The same matrix of terminal to terminal offered traffics was used each time, the traffics being a short-term, network busy hour forecast, based on recent measurements for this network. For this study, final choice routes were dimensioned to the next higher circuit (rather than to the next higher module) so that the grade-of-service on final routes would be constant irrespective of module size. The minimum size of high usage route permitted was 4 circuits, but as route sizes were required to be a multiple of the module the actual minimum route size varied from 4 to 10 circuits. After considering the aspects discussed in Section 2.3 of this paper the breakpoint used for transition from the lower preferred size to the upper preferred size was set at 60 percent of the module size.

Having computed the circuit requirements for all routes the results were used as input to another computer program which computed the costs (as annual charges) for the internal and external plant needed to meet each of the designs.

The results of this study are summarised in Table 1.

Three important findings about modular engineering of the network under consideration can be drawn from this table:

(i) any of the modules in the range 2 to 6 requires only slightly more circuits, and results in only slightly higher costs, than the reference network.

(ii) for modules in the range 2 to 6 the choice of module size does not affect to any significant degree the number of routes, the number of circuits, or the cost.

(iii) for modules greater than 6 the modular method becomes progressively less economical. This is mainly because the reduced number of first choice routes causes a higher proportion of traffic to be switched via alternative routes.

TABLE 1 EFFECT OF VARYING THE MODULE SIZE

(Note - final choice routes are NOT modular engineered)

MODULE SIZE	MINIMUM ROUTE SIZE	ROUTES REQUIRED	CIRCUITS REQUIRED	COST ($M)	COST DIFFERENCE (%)*
1	4	1091	23448	1.817	0
2	4	1091	23487	1.819	+0.11
3	6	1062	23545	1.822	+0.27
4	4	1092	23563	1.823	+0.33
5	5	1092	23591	1.824	+0.38
6	6	1062	23642	1.827	+0.55
8	8	969	24128	1.854	+2.00
10	10	883	24791	1.896	+4.35

* Compared with the reference network
(Note - total network traffic = 11796E)

Thus, from (i) there is no initial drawback to modular engineering, and from (ii), the choice of module size in the range 2 to 6 can be made on the basis of considerations other than first-up costs. For example, on routes which use carrier systems, one would favour a module size which is a multiple or sub-multiple of the group channel size. On routes using physical circuits, if all other things are equal, one might favour modules which tend to minimise the per-unit manhours required to increase the size of the route to the next higher preferred size.

A module size of 5 has been used in the Perth junction network for some time. 5 was chosen because it is a sub-multiple of all the route availabilities obtainable from an LM Ericsson ARF group selector stage. Table 1 indicates that this choice is only slightly sub-optimum for a static network. The remaining studies described in this paper use module 5 for the purposes of illustration.

The next objective of the study was to qualify the findings of Table 1 by comparing a modular engineered network with the reference network at different levels of traffic.

To simulate an actual situation in which traffic grows with time, four additional traffic matrices were produced. In these, each element was derived by increasing the equivalent element in the initial matrix by a fixed percentage. The percentages used were 5, 10, 20 and 30 percent. Thus, if for example, the network is growing as a whole at an average rate of 5 percent per annum, then the four matrices represent the equivalent of about 1, 2, 4 and 6 years growth. Naturally, these matrices would only be approximations to the real situation as there will certainly be variations in exchange growth rates, and the community of interest between exchanges is known to change with time.

The model network was dimensioned for each traffic matrix firstly with circuit increments of 1 circuit (the reference network) and then with modules of 5 circuits. Factors other than traffic changes which could influence the results (eg route availabilities and marginal occupancies) were kept constant. Final routes were dimensioned to the nearest circuit for constant grade of service. The annual charges for the internal and external plant needed to meet each design were then computed. The results of this study are shown in Table 2.

TABLE 2 COMPARISON OF THE SAME NETWORK WITH MODULES 1 AND 5

(Note - final choice routes are NOT modular engineered)

NETWORK TRAFFIC		ROUTES REQUIRED			CIRCUITS REQUIRED			COST ($M)		
		MODULE		DIFFER-	MODULE		DIFFER-	MODULE		DIFFER-
(E)	(%)	1	5	ENCE	1	5	ENCE(%)	1	5	ENCE(%)
11796	100	1091	1092	1	23448	23591	+0.61	1.817	1.824	+0.38
12387	105	1108	1110	2	24348	24494	+0.60	1.887	1.895	+0.43
12978	110	1128	1129	1	25301	25440	+0.55	1.960	1.968	+0.41
14159	120	1172	1171	-1	27109	27249	+0.52	2.100	2.109	+0.43
15340	130	1208	1207	-1	28995	29142	+0.31	2.246	2.256	+0.44

The table shows a consistent pattern with the number of circuits and costs for module 5 always at a slight disadvantage compared with the reference network. On average, module 5 requires 0.55 percent more circuits and costs 0.42 percent more than the reference network.

6. EFFECT OF CHANGING TRAFFICS ON ROUTE SIZES

The next aspect to consider was the effect of traffic growth on the size of each individual route and to do this the number of circuits on each route was compared for different pairs of traffic matrices. In this way it was possible using the traffic matrices mentioned earlier to simulate (for a network in which traffic is growing at about 5 percent per annum) the growth in route sizes from year 0 to year 1, from year 1 to year 2, from year 2 to year 4, and from year 4 to year 6 (ie cases A, B, C and D respectively in Table 3(a)). After comparing circuit quantities at the start and finish of each period, each route was categorised as being

- unchanged in size,
- decreased in size,
- increased in size, or
- a new route

The latter three categories were then lumped together as a single category "changed in size". The results of this analysis for modules 1 and 5, and minimum route sizes of 4 and 5 respectively, are shown in Table 3(a).

TABLE 3(a) CHANGES TO ROUTES WITH INCREASING NETWORK TRAFFIC

(Note - final choice routes are NOT modular engineered)

CASE	TRAFFIC (%)			ROUTES REQUIRED*		ROUTES CHANGED**		
	AT	AT	%	MODULE		MODULE		DIFFER-
	START	FINISH	CHANGE	1	5	1	5	ENCE
A	100	105	5.0	1108	1110	663	330	333
B	105	110	4.8	1128	1129	685	317	368
C	110	120	9.1	1172	1171	980	429	551
D	120	130	8.3	1208	1207	981	426	555

* At Finish
** Includes increases, decreases and new routes.

The prime advantage of modular engineering is the stabilising influence it has on the size of individual routes, and this effect is clearly shown in the Table. Over the duration of each study period module 5 permits most routes to stay fixed in size, whereas module 1 necessitates that most routes change in size. For example, in Case A (5 percent traffic growth), 60 percent of routes must change in size with module 1, but only 30 percent with module 5, and in Case C (9.1 percent traffic growth), 84 percent of routes must change in size with module 1, but only 37 percent with module 5.

In the studies presented in Table 3(a) the modular method was not used for final routes in order to ensure that the grade of service on these routes was kept at the specified design level so as not to introduce another variable into the comparison. In practice, the final routes would probably be modular engineered, but the route size would always be taken upwards to the next higher module. This slightly improves the grade of service of the final routes but increases the cost of the network. It also stabilises the size of each final route for a longer period than occurs if these routes are dimensioned to the nearest circuit. Another study was therefore conducted as in Table 3(a) but with the final choice routes modular engineered.

The increase in the number of circuits as a result of rounding the size of final routes upwards to the nearest 5 circuits comes to about 400, and as there are 200 final routes in the network, this averages 2 circuits for each final route. The increase in the number of final route circuits is equivalent to 1.7 percent of the total number of circuits in the whole network at the base traffic level (11 796E), decreasing to 1.4 percent at the 130 percent traffic level (15 340E). Costs may or may not be higher, depending on the balance between the extra cost of the junctions and the savings attributable to not having to change the size of final routes so often.

A summary is presented in Table 3(b) for modules 1 and 5 showing the stability of the network when traffic is increasing.

TABLE 3(b) CHANGES TO ROUTES WITH INCREASING NETWORK TRAFFIC

(Note - final choice routes ARE modular engineered)

CASE*	ROUTES CHANGED			CUMULATIVE TOTAL OF ROUTE CHANGES		
	MODULE		DIFFER-ENCE	MODULE		DIFFER-ENCE
	1	5		1	5	
A	663	265	398	663	265	398
B	685	238	447	1348	503	845
C	980	370	610	2328	873	1455
D	981	377	604	3309	1250	2059

* See Table 3(a) for explanation

Now, in Case A (5 percent traffic growth), only 24 percent of routes change with module 5, and in Case C (9.1 percent growth), only 32 percent of routes change with module 5, compared with 60 percent and 84 percent, respectively, for module 1. This is the major finding to come out of the study.

It might be thought that the sudden changes in overflow traffic which occur when first choice routes change in size could result in frequent changes to the size of the alternative routes. However, examination of the data used for Table 3(b) indicates that the alternative routes out of an exchange have about the same propensity to change as the first choice routes from that exchange.

7. BENEFITS FROM MODULAR ENGINEERING

It would be possible to estimate from the last column in Table 3(b) the savings which accrue through modular engineering from being able to leave routes unaltered in size for longer periods. The cost of establishing new routes and of altering the size of existing routes includes a once-only fixed component (independent of the magnitude of the change in the number of circuits). If, for example, the overall fixed cost for all participating Engineering Sections of establishing or changing a route was equivalent to 1 man-day (a conservative figure for the Perth junction network), then the savings would be 398, 447, 610 and 604 man-days for Cases A, B, C and D respectively, and the cumulative savings would be 398, 845, 1455 and 2059 man-days after 1, 2, 4 and 6 years respectively. (If a smaller module is used (eg 2 or 3) the advantage gained through stability in route sizes would be less than with a larger module (eg 4 or 5) since, on average, the routes will have to change in size more often).

When an alteration is made to the size of a working route it is usually necessary for some of the existing circuits to be made temporarily unavailable, thus reducing the traffic carrying capacity of the route. Also, there is a possibility that faults will be induced in the course of effecting the change. It follows, that by reducing the frequency of changes through the use of modular engineering, these and other disruptive effects can be reduced.

An advantage to the Drafting and Installation Sections in having a modular engineered network is that, since only certain sizes of circuit groups are permitted, a smaller number of standard interconnection patterns can be used for the grading of circuits. In addition, in the absence of any other over-riding factor, the choice of module size (or sizes) may be suited to the convenience of the Installation Section technicians. Also, under growing traffic, although the total number of circuits to be increased is not fewer under modular engineering, the number of routes to be changed is decidedly less and this factor makes the task of managing the changes considerably easier for Sections such as Planning, Drafting and Installation.

Although this paper has concentrated discussion on a network engineered with a fixed module, an appropriate combination of modules (eg. 2,3,4 and 5), selected as outlined in Section 2.2, could possibly yield greater benefits than if only one module size is used. This concept will be the subject of a later study.

The modular engineering philosophy, with minimum route size of 5 circuits and module size of 5 circuits, has been in use in the Perth metropolitan telephone network for a few years. No special problems have arisen and the system is considered by Planning, Installation, Drafting and Operations Sections to be more satisfactory than the method previously used in which circuit increments and decrements of 1 circuit were permitted.

8. EFFECT OF MODULAR ENGINEERING ON H-FACTORS

If the number of circuits specified by a modular engineering design for a high usage route is different from the exact theoretical quantity then this means departing from the correct marginal occupancy (H-factor) and is, in fact, equivalent to designing with a different H-factor. (This happens to some extent even with the traditional increment of 1 since for a given H-factor the exact circuit quantity will generally be non integer, and the quantity must be rounded (eg to the nearest integer value), thus slightly altering the H-factor.)

Table 4 illustrates the effect of rounding on H-factors for some typical route designs from an LM Ericsson ARF-10 two-stage crossbar group selector stage. For the purposes of this illustration the module size used is 5 and the breakpoint is 60 percent of the module range. The traffics have been deliberately chosen to give ideal circuit quantities at the extreme ends of each module range - eg for a rounded quantity of 10, the lower admissible requirement is 8.0 circuits, and the upper limit is 12.9 circuits. The table thus depicts the 'worst' cases.

The absolute difference between the actual and correct value of H is at its greatest (about $\pm$ .20) for the combination of low availability and low traffic, and decreases as the traffic and availability increase. The difference is less than $\pm$ .10 for offered traffics in excess of about 25 E.

Naturally, the discrepancy between the specified and actual H-factors is more pronounced the larger the module size. If it is desired to minimise the discrepancy, a range of module sizes could be used, as suggested in Section 2.2, using a module size which is roughly proportional to the size of the route.

However, it has already been noted in Section 5 that using modules in the range 2 to 6 has a negligible effect on either the number of circuits or costs of the total network. Hence, the fact that the H-factors being used are different from the ideal values is apparently of little significance to the final outcome. To further test this hypothesis the reference network was dimensioned with H-factors which differed from the ideal by up to $\pm$ 0.20. This was done for five traffic matrices with total network traffic ranging from 11 796E to 15 340E. On average, the cost of the resulting junction network was only 2.0 percent more than the cost of the equivalent reference network dimensioned with the ideal set of H-factors. This is in line with other studies which have shown that alternative routed networks are not very cost sensitive in the vicinity of the optimum design. As a corollary, it would seem that there is little point in the first place in trying to compute H-factors to great accuracy, especially if modular engineering is employed.

9. CONCLUSIONS

Levine and Wernander showed that in a once-off design (ie based on static traffic), a marginal case exists to support the use of modular engineering in alternative-routed inter-city trunk networks. The present paper has shown that for a metropolitan network employing alternative routing the first-up costs are only slightly higher with circuit modules of 2, 3, 4, 5 or 6 than with the traditional increment of 1 circuit. A second important

TABLE 4 VARIATION TO H-FACTORS

AVAIL-ABILITY	TRAFFIC OFFERED (E)	IDEAL H-FACTOR	IDEAL CIRCUIT QUANTITY	MODULAR ENGINEERED		DIFFERENCE TO H-FACTOR
				CIRCUIT QUANTITY	H-FACTOR	
5	6.2	.60	3.0	5	.40	-.20
5	7.7	.30	7.9	5	.50	.20
10	12.5	.68	8.0	10	.56	-.12
10	12.6	.41	12.9	10	.56	.15
10	19.0	.30	22.9	20	.40	.10
10	19.2	.30	23.0	25	.24	-.06
20	25.0	.40	28.0	30	.31	-.09
20	25.1	.22	32.9	30	.31	.09
20	48.0	.30	57.9	55	.37	.07
20	48.1	.30	58.0	60	.26	-.04

finding is that when one considers the dynamic situation - ie a network in which inter-exchange traffic is growing from year-to-year, a very convincing case exists for the use of modular engineering by virtue of the significant savings in administrative and installation costs which accrue on account of the fact that routes stay fixed in size over a longer period than with the traditional increment of 1. A third finding is that although the modular concept results in the use of non-ideal marginal occupancies on high usage routes, this is of little consequence.

10. REFERENCES

1. FARR, J P A Computer System for Designing Telephone Networks. Conference on Computers in Engineering, Perth, The Institution of Engineers, Australia. September 6-7, 1976. Preprint of Papers.

2. LEVINE, S W and WERNANDER, M A Modular Engineering of Trunk Groups for Traffic Requirements. Fifth International Teletraffic Congress, 1967.

3. BRIDGFORD, J N The Geometric Group Concept and its Application to the Dimensioning of Link Access Systems. Fourth International Teletraffic Congress, 1964.

4. Design and Dimensioning of Alternate Routed Networks. Telecom Australia Engineering Instruction, Planning Traffic J1300, 1965.

5. WERNANDER, M A Personal Communication. May, 1976.

Discussion

G. THIERER, Germany : *In this paper, an interesting method for modular engineering of metropolitan networks is presented. Can this method also be applied economically to long distance DDD networks.*

P. FARR, Australia : *It is true that my own studies have dealt specifically with the application of the modular engineering method to metropolitan networks. Yet Sections 1, 2, 3, 7 and 8 of my paper are quite general and are relevant to long distance networks.*

Some particular factors may be more relevant to long distance networks than to metropolitan networks. For example,

- *(i) alternative routing may not be available. Full benefits are not available in this case.*
- *(ii) Bothway routes are sometimes used in long distance networks. This does not present more of a problem with modular engineering than under conventional dimensioning.*
- *(iii) Carrier and PCM systems are extensively used in long distance networks. If modular engineering is used and the module size is made equal to the group channel size (e.g. 12), the utilisation of the transmission facility is improved since each basic design module would have no unused channels.*
- *(iv) Since a long distance trunk group may pass through many intermediate points the costs of making a change in group size may be much higher than in a metropolitan network, thus favouring the modular method. Also, less multiplexing equipment may be needed at intermediate points.*

Clearly, a special analysis is necessary for a network in order to make correct choices in regard to the options mentioned in Section 2 of my paper (i.e. minimum route size, module size, breakpoint and policy on final routes).

Levine and Wernander (Ref. 2) investigated the possible application of the modular method to long distance networks in USA, and their study method and findings will be of value to persons interested in long distance networks. The AT&T Co. has recently adopted the modular method for dimensioning long distance routes. Up to a requirement of 18 trunks, groups are sized exactly. Beyond this only modules of 12 are used.*

**Author's note : Mr. Eisenberg has since informed me that this will likely be changed to 12.*

M. EISENBERG, U.S.A. : *The same administrative savings discussed in your paper could be achieved by allowing the initial design to consist of trunk groups of any size, but to make changes only if the magnitude exceeds some minimum value. Full modular engineering can be justified only by savings resulting from granularity in switching and termination equipment. Have you obtained any evidence justifying full modular engineering rather than simply imposing a minimum threshold for making changes.*

P. FARR, Australia : *There would be many similarities between modular engineering as I have described it and a method in which a route is only allowed to change in size when it warrants an increase by a given amount. But I have not done any studies on this latter method.*

Depending on local considerations there may be good reasons to prefer true modular engineering to the alternative which you have suggested. For example, off an ARF crossbar group selector stage it is convenient to size circuit groups in modules of 5 because this is a sub-multiple of all the possible route availabilities (viz. 5, 10, 20 and 40). Again where carrier systems are utilised a module which is a multiple or sub-multiple of the basic facility unit could be strongly preferred as I said earlier in reply to the question on long distance trunks. I repeat that an analysis be made of factors which are relevant to each network.

BIOGRAPHY

JOHN PETER FARR joined the Postmaster-General's Department as a cadet engineer in 1959, and graduated with Honours in Electrical Engineering in 1963. He then completed a Master of Applied Science degree at the University of British Columbia with a thesis on the use of holograms for compressing the bandwidth required for the transmission of television signals. He also holds a degree in Economics.

In 1966 while working for the British Columbia Telephone Company he developed one of the first computer-based methods for optimising the location of tandem in metropolitan telephone networks employing alternative routing. Returning to Telecom Australia he spent some years in Planning and laid the foundation for computer utilisation in such fields as processing traffic measurements, forecasting network traffics and dimensioning junction circuit requirements on a network basis. In the last few years he has worked in the fields of exchange installation, radio, and telegraphs and data; in the latter area he was the Project Manager for the installation of the Perth centre in the Common User Data Network. Currently he is Senior Engineer (Switching and Transmission) in Metropolitan Operations and has a special involvement in the introduction of ARE-11 stored program control exchanges in Australia.

Throughout he has maintained a close involvement in the planning and optimisation of telephone networks and was responsible for the studies on modular engineering presented in this paper, and for the world's first full-scale implementation of the method in the Perth metropolitan network.

A Mathematical Model for the Long Term Planning of a Telephone Network

S. BRUYN

University of Adelaide, Adelaide, Australia

ABSTRACT

This paper presents a mathematical model for the long-term planning of a telephone junction network. The use of this model to minimize the cost of a network whilst maintaining grade of service requirements at each time period results in a large non-linear programming problem. A dynamic programming algorithm to solve this problem is presented and applied to a practical problem.

1. INTRODUCTION

This paper presents a mathematical model for the planning of a telephone junction network over a number of time periods. The model is based on a model for a single period developed by Berry [1,2]. The network has alternative routing and full availability conditions. The model results in a large mathematical programme with a non-linear, non-differentiable objective function. In the case of the Adelaide telephone network for five time periods the programme involves over 17,000 variables and over 22,000 constraints.

In order to overcome the difficulty of size, the problem was formulated as a dynamic programming problem and an algorithm which reduces the problem to a series of smaller subproblems was developed. This algorithm is based on the Progressive Optimality Algorithm developed by Howson and Sancho [4], which iteratively applies a general two-stage solution to successive overlapping stages of the problem. During the calculation of any two-stage solution, the variables for the other stages are not required and hence need not be stored in core storage. Thus the size of the problem which can be handled on a given computer is limited by the size of the two-stage problem, not by the size of the overall problem.

The two-stage problem which arises in the dynamic programming algorithm, although much smaller than the main problem, is still a large problem, with over 3,000 variables for the Adelaide network. Berry [1,3] obtained a solution to a similar problem using a modified form of Rosen's gradient projection method. In the long term planning problem however, while the constraints have the same form as in Berry's problem, the objective function is not differentiable everywhere and hence the gradient may not exist at a given point. A subgradient exists at all points and this is used if the gradient does not exist.

The algorithm has been tested using the Adelaide telephone network and the results of these tests are discussed in section 4.

2. MATHEMATICAL MODEL

2.1 BACKGROUND

An alternative routing telephone network can be represented by a directed graph with the nodes representing exchanges and the links representing the junctions between them. A pair of exchanges, one originating traffic and the other terminating traffic is called an origin-destination pair or, more concisely, an O-D pair. There are two types of links which should be distinguished. Those links which carry traffic directly between O-D pairs are called direct links. Links connecting O-D pairs via tandem exchanges carry traffic which overflows from direct links and are called overflow links. The direct links are labelled, in some order, by the integers $1,2,\ldots,k,\ldots,N$ and the overflow links by $\hat{1},\hat{2},\ldots,\hat{i},\ldots,\hat{N}$. The O-D pairs are labelled to correspond with the direct links between them. In some cases the traffic offered to an O-D pair is too small to justify provision of a direct link and when this occurs the direct link is still allowed to exist but is considered to have zero junctions.

Between each O-D pair there is a number of permissible routes or chains connecting them. These chains are denoted by $R_1^k, R_2^k, \ldots, R_{j(k)}^k$ where R_1^k is the direct link, k, and $j(k)$ is the total number of permissible chains between O-D pair k. The traffic carried on chain R_k^j, between O-D pair k is called a chain flow and is denoted by h_k^j. The total traffic offered to O-D pair k, measured in erlangs, is denoted by t^k. This traffic is first offered to the direct link R_1^k which carries h_1^k. The remainder is offered to R_2^k which carries h_2^k and so forth. The fraction of the total traffic offered to O-D pair k which overflows from the last chain is called the traffic congestion and is denoted by B^k.

Given a chain flow pattern for the whole network, the number of junctions on each link can be calculated using the following dimensioning formula developed by Berry [1,2] using an equivalent random approach.

$$n = f + A\left[\frac{(M-f)}{(M-f-1)(M-f)+v} - \frac{M}{(M-1)M+V}\right] \tag{1}$$

where M is the mean of the traffic offered to the link
V is the variance of the traffic offered to the link
f is the traffic carried on the link
v is the variance of the traffic overflowing from the link
A is the equivalent random traffic which produces overflow traffic with mean M and variance V.

The total junction cost for the network is given by

$$C = \sum_{k=1}^{N} c_k n_k + \sum_{i=1}^{\hat{N}} \hat{c}_i \hat{n}_i \tag{2}$$

where c_k is the cost per junction on direct link k.
$\hat{c}_i$ is the cost per junction on overflow link i.

The following constraints are placed on the chain flows by conservation considerations.

$$\sum_{j=1}^{j(k)} h_j^k = b^k \quad \text{for all } k \tag{3}$$

where

$$b^k = t^k(1-B^k) \tag{4}$$

$$h_j^k \geq 0 \quad \text{for all } j,k. \tag{5}$$

As the numbers of junctions, $n_k, \hat{n}_i$, can be expressed as non-linear functions of the chain flows, the problem, minimize C subject to the constraints given by (3), (4) and (5), is a non-linear programming problem with linear constraints.

A solution to this large non-linear programme was obtained by Berry [1,3] using a modified form of Rosen's gradient projection method which used the special properties of the constraint equations to obtain an explicit form for the gradient projection.

2.2 THE LONG-TERM PLANNING PROBLEM

Consider a series of time periods, denoted by $0,1,2,\ldots,p,\ldots,T$ with an existing telephone network at period $p=0$. The notation described above will be used with the addition of an extra subscript, p, denoting the time period, where necessary e.g. $h^k_{j\,p}$ is the flow carried on the jth choice route between 0-D pair k at time period p.

The cost of changing the number of junctions on a link from time period $p=0$ to the time period $p=1$ is simply the increase in number of junctions multiplied by the cost per junction at $p=1$. It is assumed that a decrease in the number of junctions results in zero cost. The incremental cost of changing the whole network from its state at $p=0$ to its state at $p=1$ is given by

$$C_1 = \sum_{k=1}^{N} c_{k\,1}[\max(n_{k\,0},n_{k\,1})-n_{k\,0}]+\sum_{i=1}^{\hat N}\hat c_{i\,1}[\max(\hat n_{i\,0},\hat n_{i\,1})-\hat n_{i\,0}] \qquad (6)$$

Similarly, from $p=1$ to $p=2$ the increment cost is

$$C_2 = \sum_{k=1}^{N} c_{k\,2}[\max(n_{k\,0},n_{k\,1},n_{k\,2})-\max(n_{k\,0},n_{k\,1})]$$
$$+\sum_{i=1}^{\hat N}\hat c_{i\,2}[\max(\hat n_{i\,0},\hat n_{i\,1},\hat n_{i\,2})-\max(\hat n_{i\,0},\hat n_{i\,1})] \qquad (7)$$

In general,

$$C_p = \sum_{k=1}^{N} c_{k\,p}[\max_{q\le p}(n_{k\,q})-\max_{q<p}(n_{k\,q})]+\sum_{i=1}^{\hat N}\hat c_{i\,p}[\max_{q\le p}(\hat n_{i\,q})-\max_{q<p}(\hat n_{i\,q})]. \qquad (8)$$

The total incremental cost over all time periods is then

$$C = \sum_{p=1}^{T} C_p. \qquad (9)$$

If at each time period, the network must satisfy grade of service requirements then the long term planning problem becomes the following mathematical programme.

$$\text{Minimize } C = \sum_{p=1}^{T} C_p \qquad (10)$$

$$\text{subject to } \sum_{j=1}^{j(k)} h^k_{j\,p} = b^k_p \quad \text{for all } k,p \qquad (11)$$

$$h^k_{j\,p} \ge 0 \quad \text{for all } j,k,p \qquad (12)$$

2.3 DYNAMIC PROGRAMMING FORMULATION

Consider a sequential decision process with input state at stage p given by vector

$$X_p = (x_{1\,p},\ldots,x_{k\,p},\ldots,x_{N\,p},x_{N+1\,p},\ldots,x_{N+i\,p},\ldots x_{N+\hat N\,p})$$

where

$$x_{k\,p} = \max_{q<p} n_{k\,q}$$

$$x_{N+i\,p} = \max_{q<p} \hat n_{i\,q}$$

The decision variables are given by the vector

$$Y_p = X_{p+1} - X_p,$$

where X_{p+1} is the output state.

The return function for stage p is

$$g_p(X_p,X_{p+1}) = (X_{p+1}-X_p)\cdot \underline{c}_p$$

where $\underline{c}_p = (c_{1\,p},\ldots,c_{N\,p},\hat c_{1\,p},\ldots,\hat c_{\hat N\,p})$

i.e. $g_p(X_p,X_{p+1}) =$

$$\sum_{k=1}^{N} c_{k\,p}[\max_{q\le p}(n_{k\,q})-\max_{q<p}(n_{k\,q})+\sum_{i=1}^{\hat N}\hat c_{i\,p}[\max_{q\le p}(\hat n_{i\,q})-\max_{q<p}(\hat n_{i\,q})]$$
$$= C_p$$

If the decision process is considered to have additive returns and the X_p lie in a bounded domain implicitly given by equations (11) and (12), then the solution of the dynamic program

$$\text{Minimize } \sum_{p=1}^{T} g_p(X_p,X_{p+1}) \qquad (13)$$

is identical to the solution of the programme given by equations (10), (11) and (12).

3. SOLUTION OF THE MATHEMATICAL PROGRAMME

3.1 A DYNAMIC PROGRAMMING ALGORITHM

The algorithm to be described iteratively solves the following two-stage problem.

$$\text{Minimize } G_p = \sum_{k=1}^{N} c_{k\,p}[\max_{q\le p} n_{k\,q}-\max_{q<p} n_{k\,q}]$$
$$+\sum_{i=1}^{\hat N}\hat c_{i\,p}[\max_{q\le p}\hat n_{i\,q}-\max_{q<p}\hat n_{i\,q}]+\sum_{k=1}^{N} c_{k\,p+1}[\max_{q\le p+1} n_{k\,q}-\max_{q<p+1} n_{k\,q}]$$
$$+\sum_{i=1}^{\hat N}\hat c_{i\,p+1}[\max_{q\le p+1}\hat n_{i\,q}-\max_{q<p+1}\hat n_{i\,q}] \qquad (14)$$

while holding $n_{k\,0},\hat n_{i\,0},\ldots,n_{k\,p-1},\hat n_{i\,p-1},n_{k\,p+1},\hat n_{i\,p+1}$ constant for all k and i

i.e. minimize with respect to $n_{k\,p}$ and $\hat n_{i\,p}$.

$$\text{subject to } \sum_{j=1}^{j(k)} h^k_{j\,p} = b^k_p \quad \text{for all } k \qquad (15)$$

$$h^k_{j\,p} \ge 0 \quad \text{for all } j,k \qquad (16)$$

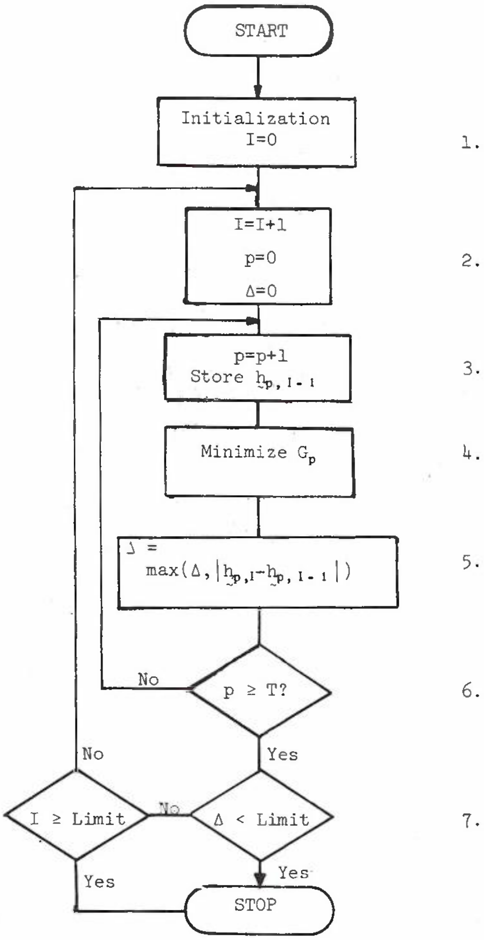
The algorithm is started by assigning an initial chain flow pattern to each time period and proceeds as illustrated in Fig.1.

This algorithm is based largely on the Progressive Optimality Algorithm described by Howson and Sancho [4]. There is, however one major difference between the algorithms and that is in the two-stage problem which is solved in each iteration. The Progressive Optimality Algorithm requires that the initial and terminal states of the two-stage problem, i.e. $\max_{q<p}\hat n_{k\,q}$, $\max_{q<p}\hat n_{i\,q}$ and $\max_{q\le p+1} n_{k\,q}$, $\max_{q\le p+1}\hat n_{i\,q}$, remain fixed. Forcing $\max_{q\le p+1} n_{k\,q}$ and $\max_{q\le p+1}\hat n_{i\,q}$ to remain constant would add non-linear constraints to the problem and hence make it much more difficult to solve. The numbers of junctions at period p+1, i.e. $n_{k\,p+1}$ and $\hat n_{i\,p+1}$, can be kept constant without affecting the constraints because the chain flows at one period are independent of those at another period. This difference from the Progressive Optimality Algorithm means that Howson and Sancho's proof of convergence of the algorithm is not valid in this case even if the stage return function g_p is differentiable. A proof of convergence has not yet been developed however computational evidence, which is described in the next section, supports the validity of the algorithm.

The algorithm requires the terminal state to be fixed and to achieve this an artificial terminal state is added. This extra stage has all link costs set to zero so as to have no effect on the total cost and to ensure that $\max_{q\le T+1} n_{k\,q}$ and $\max_{q\le T+1}\hat n_{i\,q}$ do not vary. $n_{k\,T+1}$ and $\hat n_{i\,T+1}$ are set at very large values.

3.2 A SOLUTION OF THE TWO-STAGE PROBLEM

The mathematical programme given by equations (14), (15) and (16) has constraints of a form identical to those of the mathematical programme given by equations (2), (3) and (6). This suggests the use of a similar method to that described by Berry [1,3]. However, the gradient projection method requires the objective function to have continuous first derivatives and equation (14) is not differentiable with respect to the chain flow variables at period p, when $n_{k\,p}$ is equal to



1. Set initial values of the chain flows.
   I is a control count of the number of iterations.

2. Start a new iteration at stage 0.
   Delta is a control variable.

3. Store $h_{p,n-1}$ (chain flow pattern for the network at period p after the n-1th iteration) until calculation of delta below.

4. Solve the two-stage problem.

5. Calculate delta.

6. Advance another stage until final stage is reached.

7. If delta is less than the accuracy limit or if the maximum number of iterations have been completed then STOP.

Fig. 1. Dynamic Programming Algorithm

$\max_{q<p} n_{kq}$ or n_{kp+1}, similarly for $\hat{n}_{kp}$. Considering just one link, the following possibilities arise,

1. $\max_{q<p} n_{kq} < \max_{q\le p} n_{kq} < \max_{q\le p+1} n_{kq}$

2. $\max_{q<p} n_{kq} = \max_{q\le p} n_{kq} < \max_{q\le p+1} n_{kq}$

3. $\max_{q<p} n_{kq} < \max_{q\le p} n_{kq} = \max_{q\le p+1} n_{kq}$

4. $\max_{q<p} n_{kq} = \max_{q\le p} n_{kq} = \max_{q\le p+1} n_{kq}$

Case 1. The two-stage cost associated with the link is

$$G_{kp} = (c_{kp} - c_{kp+1})n_{kp} + c_{kp+1}n_{kp+1} - c_{kp}\max_{q<p} n_{kq}$$

and the general partial derivative is

$$\frac{\partial G_{kp}}{\partial h^{\ell}_{jp}} = (c_{kp} - c_{kp+1})\frac{\partial n_{kq}}{\partial h^{\ell}_{jp}}$$

Case 2. The two-stage cost associated with the link is

$$G_{kp} = c_{kp+1}n_{kp+1} - c_{kp+1}\max_{q<p} n_{kq}$$

If $n_{kp} < \max_{q<p} n_{kq}$, then the general partial derivative is

$$\frac{\partial G_{kp}}{\partial h^{\ell}_{jp}} = 0$$

If $n_{kp} = \max_{q<p} n_{kq}$, then the general partial derivative does not exist.

Case 3. The two-stage cost associated with the link is

$$G_{kp} = c_{kp}n_{kp} - c_{kp}\max_{q<p} n_{kq}$$

If $n_{kp} > n_{kp+1}$, then the general partial derivative is

$$\frac{\partial G_{kp}}{\partial h^{\ell}_{jp}} = c_{kp}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}$$

If $n_{kp} = n_{kp+1}$ then the general partial derivative does not exist.

Case 4. The two-stage cost associated with the link is

$$G_{kp} = 0$$

If $n_{kp} < \max_{q<p} n_{kq}$ then the general partial derivative is

$$\frac{\partial G_{kp}}{\partial h^{\ell}_{jp}} = 0$$

If $n_{kp} = \max_{q<p} n_{kq}$ then the general partial derivative does not exist.

A vector $\underset{\sim}{v}$ is a subgradient of a convex function f at a point $\underset{\sim}{z}$ if

$$f(\underset{\sim}{x}) \ge f(\underset{\sim}{z}) + \underset{\sim}{v}^T(\underset{\sim}{x} - \underset{\sim}{z}) \quad \text{for all } \underset{\sim}{x}$$

If f is differentiable and finite at $\underset{\sim}{z}$ then $\underset{\sim}{v}$ is unique and is the gradient, $\nabla f(\underset{\sim}{z})$, of f at $\underset{\sim}{z}$. More detailed information about the theory of subgradients can be found in Rockafellar [5].

The optimization of the two-stage problem is carried out using the gradient projection method which is modified by using a subgradient when the gradient does not exist. The subgradients, of the various components of the objective function, which are used are the partial derivatives, where they exist and elsewhere the following,

Case 2. If $n_{kp} = \max_{q<p} n_{kq}$, then the subgradient is

$$\gamma_k = c_{kp}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}\cdot\alpha - c_{kp+1}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}\cdot\alpha \qquad 0 \le \alpha \le 1$$

Case 3. If $n_{kp} = n_{kp+1}$, then the subgradient is

$$\gamma_k = c_{kp}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}} - c_{kp+1}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}\cdot\alpha \qquad 0 \le \alpha \le 1$$

Case 4. If $n_{kp} = \max_{q<p} n_{kq}$ then the subgradient is

$$\gamma_k = c_{kp}\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}\cdot\alpha \qquad 0 \le \alpha \le 1$$

The partial derivatives, $\frac{\partial n_{kp}}{\partial h^{\ell}_{jp}}$, are calculated, with some slight modifications, as in Berry [1].

4. COMPUTATIONAL RESULTS

The mathematical program was applied to the Adelaide telephone network. Traffic dispersions and junction costs were supplied by the Australian Telecommunications Commission in Adelaide for 5 time periods, viz. 1975, 1980, 1985, 1990, 1995. The junction costs for 1975 were based on capital costs and those for 1980, 1985, 1990, 1995 were calculated by discounting the 1975 costs by factors of 0.650, 0.275, 0.075, 0.013, respectively. The network consists of 32 exchanges which both originate and terminate traffic, 5 exchanges which terminate traffic only, and 4 tandem exchanges. There are 1141 O-D pairs with each pair having 2 overflow routes and, unless the offered traffic is below an arbitrary cut-off figure (1.85 erlangs in 1980), a direct route is also provided.

In order to obtain a fixed initial state for the algorithm, the network was optimized for 1975 using the method described by Berry [1,3]. The algorithm was tested with 3 different starting points. The first starting point was obtained by optimizing the network for each time period as was done for 1975. The second starting point was obtained by assigning the chain flows so that for each O-D pair, at each time period, the proportion of flow on each route was the same as in 1975. The third point was obtained by assigning traffic for each O-D pair in the ratio to offered traffic of .6, .2, .18 to the first, second and third choice routes respectively, or, if no direct link is provided, in the ratio .6 and .38. In all cases the congestion for every O-D pair was set at 0.02. The main algorithm was limited to 5 complete iterations and the algorithm for finding the solution of the two-stage problem was limited to 100 iterations. The main algorithm was terminated if, after any complete iteration, the maximum change in any chain flow variable was less than 0.5 erlangs. The two-stage algorithm was terminated if the two-stage cost decreased by less than one dollar in successive iterations.

Initial Point	Initial Cost	Final Cost	c.p.time (sec)
1	$1,336,662	$1,326,658	13,610
2	$1,340,740	$1,325,052	8,085
3	$1,986,960	$1,409,105	12,276

Table 1.

Table 1 shows the decrease in cost for each starting point after 5 iterations and the central processor time taken on a CDC 6400 computer. The time taken for the first starting point includes the time taken to set up the starting point. The number of junctions on each link was rounded up to the nearest integer and a count was taken of the number of links which had a decrease in the number of junctions between successive time periods. This gives the number of links which have unused junctions at any time period. These numbers were obtained for each starting point and their respective final points obtained after 5 iterations of the algorithm, and are given in Table 2.

Initial Point	No.of links with decrease in junctions	
	Initial Point	Final Point
1	31	5
2	13	10
3	229	82

Table 2.

Table 3 gives the total cost of the network at each time period for the first and second starting points at both the initial and final points of the algorithm.

Year	First Starting Point		Second Starting Point	
	Initial	Final	Initial	Final
1980	2,894,588	2,887,102	2,895,958	2,885,172
1985	1,591,185	1,586,457	1,594,815	1,585,242
1990	532,693	531,299	535,449	531,288
1995	108,960	108,667	109,7907	108,937

Table 3.

These results show that as well as decreasing the cost of the long-term problem, the algorithm also decreases the total cost at each time period. This suggests that, for the Adelaide network at least, a near optimal solution to the long-term problem will result in chain flow patterns which give near optimal solutions to the single period problems.

The traffic dispersion forecasts used in these tests have the property that the offered traffic increases with the time period for almost every O-D pair. In order to test the algorithm with data which did not have this structure, traffic dispersion figures were obtained by using a set of random numbers, uniformly distributed between 0 and 40. The algorithm still gave a reduction in cost at every iteration and after 5 iterations the cost had been reduced from $5,107,530 to $4,486,487. This shows that the algorithm is not dependent on link junction numbers increasing with time and can deal with problems having a large number of links where $\max_{q<p} n_{kq} = \max_{q<p} n_{kq}$.

5. CONCLUSION

A mathematical model for the long-term planning of a telephone junction network and an algorithm for obtaining the minimum cost for the network have been developed. The algorithm solves the problems of having a large non-linear mathematical programme and a non-differentiable objective function. Tests on a practical network have shown that the algorithm does reduce the cost of the network but that it needs a large amount of computer processing time. Further work needs to be done in an attempt to accelerate the convergence of the algorithm and hence reduce the processing time.

REFERENCES

1. L.T.M. BERRY "A Mathematical Model for Optimizing Telephone Networks" Thesis for PH.D. (Adelaide) December 1971.
2. L.T.M. BERRY "An Explicit Formula for Dimensioning Links Offered Overflow Traffic", A.T.R., Vol.8, No.1, (1974).
3. L.T.M. BERRY "On the Solution of a Structured Non-linear Programme" To be published J.Austral. Math. Soc. (Series B)
4. H.R. HOWSON & N.G.F. SANCHO "A New Algorithm for the Solution of Multi-State Dynamic Programming Problems",8, pp 104-116, (1975)
5. R.T. ROCKAFELLAR "Convex Analysis", Princeton University Press (1970).

Discussion

D. MANFIELD, Australia : *In your paper the variables of the objective function are the numbers of links and the constraints are expressed in terms of the chain flows. Could Mr. Bruyn give us some idea of how the cost function was obtained in terms of the numbers of links from the chain flows generated by the algorithm and whether these numbers were generated at each iteration.*

S.J. BRUYN, Australia : *Junction numbers can be calculated from the flow on the link using Berry's dimensioning formula. The objective function is considered as a non-linear function of the chain flows and is calculated directly from the chain flows when required. See references 1-3 in the paper.*

N.W. MACFADYEN, U.K. : *Could you give an indication of the sensitivity of the algorithm to errors in the traffic data and the discount factors employed.*

S.J. BRUYN, Australia : *The algorithm appears to reduce the total network cost regardless of the discount factors used, it even reduces the cost if junction costs increase with time.*

Sensitivity to changes in offered traffic has been tested by uniformly increasing all traffic by a given percentage. The percentage rise in total network cost is smaller than the rise in the traffic, e.g. 5% rise in traffic results in approximately 3% rise in cost, and 25% rise in traffic results in approximately 19% rise in cost.

D.J. SUTTON, Australia : *The junction costs used in your computational results decrease very rapidly with time. Can you explain why you chose these discount factors.*

S.J. BRUYN, Australia : *This was an error in the data used in the early experiments with the algorithm. The discount figure should have been 9% per annum which would give discount factors of .650, .422, .275, .178. If these changes are made as well as some imrpovements to computational aspects of the algorithm Tables 1, 2, 3 in the paper become:*

Initial Point	Initial Cost	Final Cost	CP Time (sec)
1	1,921,912	1,918,308	7,297
2	1,945,784	1,938,017	1,964

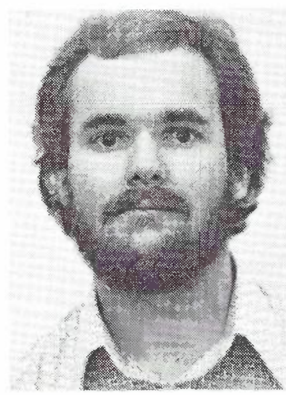
TABLE 1

Initial Point	No. of Links with Decrease in Junctions	
	Initial Point	Final Point
1	52	33
2	13	6

TABLE 2

YEAR	First Starting Point		Second Starting Point	
	Initial	Final	Initial	Final
1980	2,879,913	2,880,015	2,890,270	2,890,225
1985	2,431,711	2,431,760	2,453,745	2,453,582
1990	1,927,420	1,927,517	1,951,895	1,950,581
1995	1,436,770	1,436,763	1,458,990	1,450,995

TABLE 3



BIOGRAPHY

STEWART BRUYN completed his Bachelor of Mathematics at the University of Newcastle in 1973. Since then he has been enrolled as a Ph.D. student in the Applied Mathematics Department of the University of Adelaide under the supervision of Dr. L.T.M. Berry and Prof. R.B. Potts. His research interest is in the mathematical aspects of the long-term planning of a telephone network.

Cost Effectiveness of Traffic Measurements

Y. M. CHIN

Telecom Australia, Sydney, Australia

ABSTRACT

This paper is concerned with the optimisation of the method, where the data collected from a previous measurement plays an important part in the measurement-dimensioning cycle.

As any traffic measurement is only an estimate of the offered traffic, further usage of the data will result in errors, directly caused by the sampling process. If the sampling variance is known, this imprecision can be allowed for in the dimensioning process. This adjustment represents an additional cost penalty attributed to the practicality of traffic measurements.

In this paper, it is assumed that the cost of the traffic study is linearly related to the duration of the measurement. The cost penalty is shown to be approximately inversely proportional to the time spent in conducting the experiment. If we consider the measurement phase as part of an investment, clearly, by choosing the traffic measuring parameters, the return on investment can be maximised.

The above concept is applied to the occupancy measurement of a first choice high usage route in a simple triangular alternative routing pattern. The traditional network cost minimisation technique and full availability working are assumed.

Results obtained indicate that, from some route parameters, one can estimate the optimum time duration for a traffic study, such that the measurement costs equal the cost penalty. In general, optimum occurs where:

$$\frac{d(\text{measurement cost})}{dt} = \frac{-d(\text{cost penalty})}{dt}$$

1. INTRODUCTION

Any traffic measurement is an estimate of the average traffic flow. In principle, for a given technique, it is possible to estimate the sampling variance and other statistical parameters related to the measurement. Calculations based on the measurement will therefore be imprecise. In practice, this error is allowed for by an adjustment to the dimensioning process. This adjustment is a cost penalty arising from the practicality of traffic measurements. Generally for a given technique, doubling the duration of measurement halves the sampling variance and roughly doubles the measurement costs. In return, the reduced error decreases the cost penalty by allowing more accurate prediction of circuit requirements.

If the policy of a measurement-upgrading technique is adopted, then the measurement must be considered as part of the capital investment which can be optimised, such that no more effort should be spent in data acquisition than necessary.

For simplicity, this concept is applied to the measurement of a high usage route in a simple triangular alternative routing pattern. The traditional network cost minimisation and full availability with direct access are assumed.

Results obtained thus far indicate that given some route parameters it is possible to estimate the optimum time of a traffic study such that the return on investment for a measurement-dimensioning cycle can be maximised.

2. RESUMÉ OF SOME BASIC TOOLS

Consider the simplest triangular alternate routing pattern

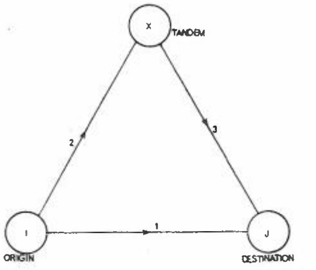


FIG.1: A SIMPLE ALTERNATE ROUTING PATTERN

The assumptions, definitions and notations used in Pratt's paper will be retained in this discussion (Ref.1). Some of the related definitions used here are as follows:-

Consider a route of N circuits, offered traffic A and carried traffic Y = (A-a), where a is the traffic lost or overflowing.

The route congestion is $G = a/A$ (1)

(i) Marginal occupancy, $H = \left(\frac{\partial Y}{\partial N}\right)_A$ (2)

$$-H = \left(\frac{\partial a}{\partial N}\right)_A$$

(ii) Marginal capacity, $\beta = \left(\frac{\partial A}{\partial N}\right)_E$ (3)

For the link i,

N_i = number of circuits provided for I-J traffic

C_i = Cost per circuit

A_i = Traffic offered

a_i = Traffic overflowing

For the simple case of Fig.1, the traditional optimisation equation is,

$$\frac{C_1}{H_1} = \frac{C_2}{\beta_2} + \frac{C_3}{\beta_3} \tag{4}$$

Further, for a full availability direct access route, a simple practical relation between A,N and H can be shown to be:-

$$H = A\ (E_N\ (A) - E_{N+1}\ (A)\ ) \tag{5}$$

where $E_N(A) = \dfrac{A^N/N!}{\sum_{i=0}^{N} A^i/i!}$, the Erlang Loss Formula. (6)

Note that H is also referred to as the Cost Factor.

3. DIMENSIONING COST PENALTY

Consider a traffic study conducted to determine the traffic on the high usage route. If the measurement period is sufficiently long then the standard deviation of the error, S, on the measured traffic Y, will be negligibly small. Subsequent dimensioning using the traditional optimisation method will result in a near "perfect" solution, such that the cost of this network is at a minimum.

If the direct route is dimensioned on the estimated mean offered traffic A with an appreciable error S, then the overflow to the final route will be different from that calculated by an amount (see appendix A).

$$\therefore \quad \Delta U = \frac{U(A+S) + U(A-S) - 2U(A)}{2} \tag{7}$$

where $U(A) = A.E_N(A)$, the overflow traffic for N ccts

$$\text{and} \quad U(A+S) = (A+S).E_N(A+S) \tag{8}$$

$$U(A-S) = (A-S).E_N(A-S)$$

The expected cost of this extra unpredicted traffic is $(\frac{\Delta U}{H})C_d$

Where C_d is the cost of the direct circuit.

This cost represents the dimensioning cost penalty in not being able to estimate the offered traffic accurately.

As a matter of convenience we shall introduce two definitions:-

$$\text{Cost Penalty} \qquad Pc = (\frac{\Delta U}{H})C_d \tag{9}$$

$$\text{Penalty Function} \qquad P = (\frac{\Delta U}{H}) \tag{10}$$

For a given A and H, the optimum number of direct circuits, N can be determined, and hence ΔU, provided that S, the sample standard deviation is also known. For a given measurement technique, S is directly related to the measurement duration, t_m.

Clearly, for a given A and H, the penalty function is related to t_m.

4. SIMPLIFIED RELATIONSHIP BETWEEN S AND t_m

For a scanning method for a route of infinite size, Hay- - ward (Ref.3) showed that the variance of n as an estimator of Y is given by -

$$V(n;Y) = \frac{M}{T}(r \cotanh(r/2) - 2) + \frac{2Y}{T^2}(T-1+e^{-T}) \tag{11}$$

Where r = ratio of scan interval to holding time
T = length of observation period
M = the unbiassed estimate of Y, the carried traffic.

$$\text{Cotanh}(r/2) = (1+e^{-r}) / (1-e^{-r})$$

$$r \text{ Cotanh}(r/2) = 2 + \frac{r^2}{6} - \frac{r^4}{360} + \frac{r^6}{15120} \ldots$$

For practical cases, $e^{-T} \to 0$, and M and Y are nearly equal, and further assume that the offered traffic A is very close to Y.

$$\text{Hence } V(n;A) \doteqdot \frac{A}{T}(2 - \frac{2}{T} + \frac{r^6}{r}) \tag{12}$$

$$\text{Now} \qquad S^2 = V(n;A)$$

and if r = 1 (eg. 3 min. scan interval, 3 min. holding time)

$$\text{then} \quad T = \frac{7A}{6S^2} \pm \frac{\sqrt{1.36A^2 - S^2A}}{S^2} \tag{13}$$

Normally, $\frac{S}{A} \leqslant \frac{1}{10}$

i.e. specifying that no more than 10% error in A

$$\text{Then} \quad T = \frac{2.33A}{S^2} \tag{14}$$

$$\text{For simplicity} \quad T = \frac{2A}{S^2} \tag{15}$$

Despite the simplifications and the assumptions made in Hayward's paper, this simple formula is sufficient for practical uses.

Given the holding time h, the observation period in real time is given by

$$t_m = \frac{2Ah}{S^2} \tag{16}$$

For our purposes we shall take h as 3 minutes and A as the average time consistent busy hour traffic. A common practice used in determining this traffic is to perform the measurement in units of days, and thence select the time consistent busy hour from the samples. Clearly, unless there is some other means of obtaining prior information as regards the busy hour, this approach of data collection means that every 60 minutes of t_m in fact represent a day of observation.

5. RELATION BETWEEN THE PENALTY FUNCTION AND MEASUREMENT TIME

For a given traffic, cost factor and standard deviation, it is possible to work out the Penalty Function P, and the length of the observation time t_m.

Numerical analysis reveals that, in general an equation of the type

$$P = \frac{1}{\alpha + Bt_m} \tag{17}$$

consistently gives a remarkably good fit to the data generated from the above system of equations. Table 1 is a list of the coefficients α, B for some typical H and A values.

Traffic, A in Erlangs	α	B	Cost Factor H	Optimum Circuit
2	0.0763 -0.3132 -1.8	7.50 22.73 81.00	0.2 0.4 0.6	4 2 1
20	0.0279 0.0238 -0.0249	3.05 5.72 11.52	0.2 0.4 0.6	27 23 20
100	0.0106 0.0149 -0.0008 -0.0714	1.489 2.622 5.319 15.101	0.2 0.4 0.6 0.8	116 108 100 89
200	0.0068 0.0110 0.0036	1.0526 1.8664 3.6410	0.2 0.4 0.6	222 212 201

Table 1: VALUES OF α, B FOR TYPICAL VALUES OF TRAFFIC AND COST FACTOR

α may be ignored, and a simple relation between the Penalty Function and measurement time (in days) results.

$$P = \frac{1}{Bt_m} \tag{18}$$

It follows that the Dimensioning Cost Penalty is:

$$P_c = \frac{C_d}{Bt_m} \tag{19}$$

6. MEASUREMENT COSTS

For a given t_m equation (19) provides a convenient means of estimating the dimensioning cost penalty due to insufficient information regarding the true traffic. In principle, it is also possible to estimate the cost of acquiring this information.

In general, measurement cost is a complicated function. For most modern mechanised measurement systems, the following costs should be taken into account.

- Development of measurement equipment
- Capital investment in the measurement equipment
- Development and maintenance of processing software
- Measurement set up and check out
- Supervision of measurement
- Analysis of reports

As an approximation, assume that a linear relationship exists between the cost of measurement, M_c, and the time spent, t_m.

$$M_c = C_m t_m \tag{20}$$

where C_m = cost of measurement/group/day.

7. OPTIMUM MEASUREMENT DURATION t_m opt

The total cost D, is a sum of the Dimensioning Cost Penalty P_c, and the Measurement Costs M_c : (see Fig.2)

$$D = \left(\frac{C_d}{Bt_m}\right) + C_m t_m \tag{21}$$

Now $$\left(\frac{dD}{dt_m}\right) = -\left(\frac{C_d}{Bt_m^2}\right) + C_m \tag{22}$$

At $$\left(\frac{dD}{dt_m}\right) = 0$$

$$\left(\frac{C_d}{Bt_m}\right) = C_m t_m \tag{23}$$

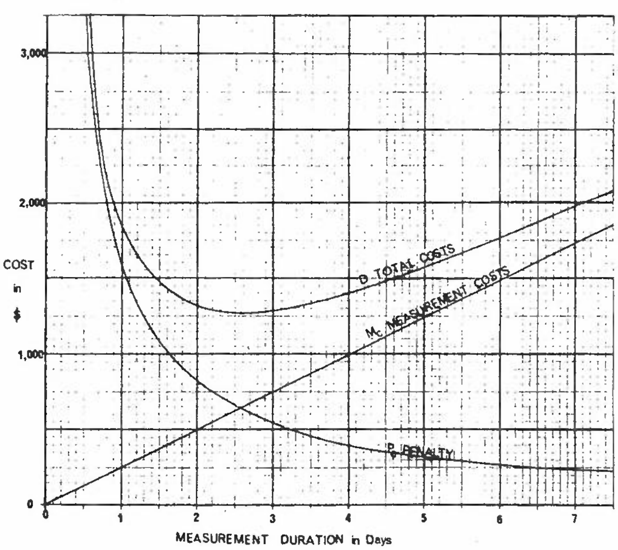


Fig.2: PENALTY, MEASUREMENT AND TOTAL COST FUNCTIONS

Hence, the optimum duration of an occupancy traffic study is:

$$t_m \text{ opt} = \frac{1}{\sqrt{BM}} \tag{24}$$

where $$M = \frac{Cm}{Cd} \tag{25}$$

In principle, M can be estimated providing the cost of a direct circuit C_d, and the cost of measurement/group/day, C_m is known. For a fair comparison, C_m and C_d must be at the same scale. If C_d represents the annual charges, C_m must also be annual charges.

8. ESTIMATING THE B CONSTANT

The optimum duration of a traffic study t_m opt is completely defined by B and M. With M estimated, the next task is to find the B constant.

Now $B = f(H,A,N)$

Numerical analysis and a graphical plot of B versus H (Fig.3) reveals that in general, for a particular N, (uniquely determined by A and a range of H) B is a linear function of H for ranges of H.

$$B = \mu H \tag{26}$$

The simplicity of equation 26 is useful for practical uses. If μ is also known, then the problem of estimating t_m opt becomes easy.

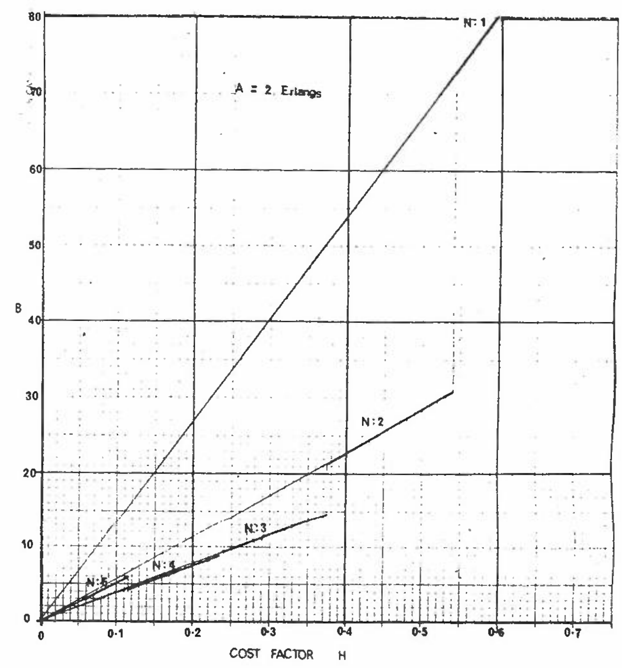


Fig. 3: TYPICAL PLOT OF B VERSUS H FOR CONSTANT TRAFFIC A

From Fig.3, one is led to suspect that μ, the gradient is some function of N (and indirectly, related to A and H).

Unfortunately, further numerical analysis of μ variation on A or H did not reveal any simple relation.

Fig.4 is a graphical representation of μ versus N for different values of traffic. Clearly, as will be shown by an example, this graph can be used to estimate the optimum measurement time period, provided other parameters are available.

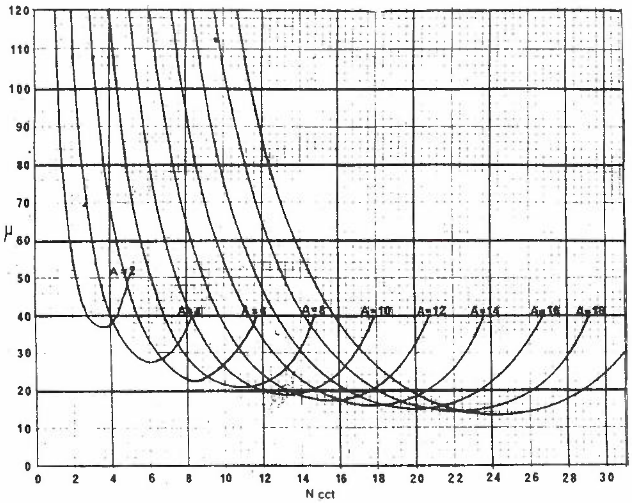


Fig.4: PLOT OF μ v N

9. ESTIMATING THE OPTIMUM MEASUREMENT TIME: AN EXAMPLE

Suppose we wish to determine the optimum observation time for a direct circuit, given the following information.

(a)	Cost of direct circuit (per annum)	Cd	= $1500
(b)	Cost of measurement/group/day (per annum)	Cm	= $ 5
(c)	Estimated traffic expected on route	A	= 10Erlang
(d)	Cost factor to be used in dimensioning	H	= 0.6

For A = 10 Erlang and at H = 0.6, the optimum direct route size given by equation (5) or from standard graphs (Refs.4, 6) is N = 9. (see Table 2)

From Fig.4, for N = 9 and A = 10, μ = 35.

The B constant as given by B = μ H is B = 35 x 0.6 = 21.

The measurement cost/circuit cost ratio M = (Cm/Cd)=0.0033.

Hence the optimum study duration, $t_m \text{ opt} = \frac{1}{\sqrt{MB}}$
$= 4$ days

Clearly, for a 5-day business cycle type of route, a minimum of 5 days will be spent on the actual measurement.

10. SENSITIVITY OF M AND B TO ERROR

The key equation in estimating the optimum measurement time is

$$t_m \text{ opt} = \frac{1}{\sqrt{MB}}$$

For a fixed M value, it is interesting to find the variation of t_m opt with respect to B and M (Table 2)

Traffic,A (Erlang)	Optimum CCt.N (H=0.6)	μ	B=μH	$t_m \text{ opt} = \frac{1}{\sqrt{MB}}$ (Days) M=0.0033	M=0.0104	M=0.033
4	3	62	37.2	2.85	1.60	0.90
6	5	46	27.6	3.31	1.86	1.05
8	7	38	22.8	3.65	2.05	1.15
10	9	33	19.8	3.91	2.20	1.24
12	11	30	18.0	4.10	2.31	1.30
14	13	27	16.2	4.32	2.43	1.37
16	15	25	15.0	4.49	2.53	1.42
18	17	23	13.8	4.69	2.64	1.48
20	20	16	9.6	5.62	3.16	1.78

TABLE 2: VARIATION OF t_m opt w.r.t. B,M.

The restriction placed on using this method is that an initial guess of the traffic must be available (possibly from past measurements). Table 2 shows the error made in estimating t_m opt is marginal. On the other hand, errors in estimating M will have a marked effect on t_m opt. Clearly for practical usage, M should be estimated more precisely.

11. CONCLUSIONS

This study has shown that from some route parameters and an initial estimate of the traffic, it is possible to estimate the optimum time for a traffic measurement.

Although the study has been confined to the case of a full availability first choice route in a simple alternative plan, this principle can be extended to other high usage routes of more complex overflow patterns. Tests (not listed in this paper) indicate that the concept can indeed be applied to first choice routes with link access and limited availability.

The application of the method requires some prior knowledge of the traffic, so that the optimum number of direct circuits, and the parameter B can be found, and the optimum study period t_m opt estimated.

Inspection of the key equation $t_m \text{ opt} = \frac{1}{\sqrt{MB}}$ and the associated graphs show that in practice, t_m opt is more sensitive to M than A. Hence, even if the initial traffic estimate is not precise (making an error in B) the error in t_m opt is relatively small.

Effort should be spent in getting a more precise estimate of the measurement-circuit cost ratio, M. As in all engineering economics, the estimation of costs represent a real problem. In reality, M is a comparison between capital costs and labour costs. The effectiveness of the measurement-provisioning policy depends on how accurately one can estimate the ratio of manhour to equipment cost.

12. ACKNOWLEDGEMENTS

The author wishes to express his thanks to:

Mr. A.H. Freeman for providing the basic concept,
Mr. R.E. Warfield for the many hours of discussions,
Mr. R.L. Edmunds and Mr. E.K. Southworth for the support and encouragements, and
the Staff of Traffic Engineering Section, New South Wales for the assistance.

APPENDIX A:- ESTIMATING THE UNPREDICTED OVERFLOW : ΔU

If an element of A Erlang is estimated (measured) with a sampling error (standard deviation) of S Erlang, then the probability density function of the estimate is given by

$$p(x) = f\left(\frac{x-A}{S}\right) \qquad \text{(A-1)}$$

normalised by letting

$$B = \left(\frac{x-A}{S}\right), \text{ we have}$$

$$p(x) = \phi(B), \qquad \text{(A-2)}$$

Where ϕ is the probability density function for a normal distribution, with mean = 0 and variance = 1.

Assume this traffic, A is offered to a route, and it is desired to estimate some quantity such as carried traffic, lost traffic, variance of lost traffic etc.

The estimate of this quantity is a function of the estimated traffic g(x), with an expected value of

$$E(g) = \int_{-\infty}^{\infty} p(x) \cdot g(x)\, dx \qquad \text{(A-3)}$$

Let $G(B) = g(x)$,

and assume G(B) can be expanded as a series

$$G(B) = \alpha_0 + \alpha_1 B + \alpha_2 B^2 + \ldots\ldots \qquad \text{(A-4)}$$

$$\text{Then } E(g) = \int_{-\infty}^{\infty} \phi(B) . (\alpha_0 + \alpha_1 B + \alpha_2 B^2 + \ldots\ldots) dB \qquad \text{(A-5)}$$

$$= \alpha_0 + \alpha_1 E(B) + \alpha_2 E(B^2) + \alpha_3 E(B^3) + \alpha_4 E(B^4) + ..$$

$$= \alpha_0 + 1.\alpha_2 + 1.3.\alpha_4 + 1.3.5.\alpha_6 + \ldots\ldots \text{(A-6)}$$

(See Ref. 5 page No.147)

$$\text{Now } G(1) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \ldots \qquad \text{(A-7)}$$

$$G(-1) = \alpha_0 - \alpha_1 + \alpha_2 - \alpha_3 + \alpha_4 - \ldots$$

$$\text{Hence } \frac{G(1) + G(-1)}{2} = \alpha_0 + \alpha_2 + \alpha_4 + \alpha_6 \ .. \qquad \text{(A-8)}$$

If the 4th and higher powers are negligible, then

$$E(g) = \frac{G(1) + G(-1)}{2} \qquad \text{(A-9)}$$

$$= \frac{g(A+S) + g(A-S)}{2}$$

This estimate differs from the value which would apply if the traffic was known precisely to be A. The expected difference is:-

$$E(d) = \frac{g(A+S) + g(A-S) - 2g(A)}{2} \qquad \text{(A-10)}$$

If $g(A) = A\,En(A)$, the overflow traffic, then

$E(d) = \Delta U$, the unpredicted overflow.

REFERENCES:

1. Pratt, C.W., The concept of Marginal Overflow in Alternate Routing. Australian Telecommunication Research 1(1967): 1&2, pp76-82.

2. Pratt, C.W., A course in Telephone Traffic Engineering. Fundamental Planning, Headquarters, Australian Post Office, Melbourne 1967 pp78.

3. Hayward, W.S., The Reliability of Telephone Traffic Load Measurements by Switch Counts. B.S.T.J. 31, March 1952, pp357-377.

4. A.P.O., Alternate Routing in Step-by-Step Networks. Australian Post Office Engineering Instruction, Planning Traffic J1150. Issue 1, 1966.

5. Papoulis, A., Probability, Random Variables and Stochastic Process McGraw Hill, Kogakusha 1965, pp147.

6. Jensen, A., Moe's Principle. The Copenhagen Telephone Company, Copenhagen 1950, Table IV.

Discussion

L. LEE, Canada : *The example in Mr. Chin's paper shows that optimum study duration is 4 days. There is an example in Mr. Barnes' paper using 20 days. Now, after their excellent papers have been presented, I wonder if Mr. Chin would like to change his mind to use an example to show 20 days as the optimum study duration, and Mr. Barnes would like to change his mind to use a 4 days example instead.*

Y.M. CHIN, U.K. : *The optimum solution listed in my paper is for a particular instance. Moreover the concept is applicable to the dimensioning/measurement of a high usage route in a triangular alternative network. Barnes' discussion is on a different application and not necessarily based on an alternative routing system. The value of 20 days was chosen for a different reason.*

R. LAUFENBURGER, U.S.A. : *Considering the cost items identified on page 3, the possibility exists of the study cost being a decreasing function with respect to duration. Have you considered the effect of such likely possibility (linear or exponential) on your conclusions.*

Y.M. CHIN, U.K. : *Although we have not considered other types of cost functions, it is entirely feasible. For instance, we could have used an equation of type*

$$Mc = C_m t_m + I,$$

where I could be the initial set up costs. In this case, D, the total cost is:

$$D = \left(\frac{Cd}{Bt_m}\right) + C_m t_m + I$$

$$\text{At } \left(\frac{dD}{dt_m}\right) = 0, \quad \text{we arrive at}$$

$$t_m \text{ opt} = \frac{1}{\sqrt{BM}}$$

L. LEE, Canada : *To estimate the optimum measurement time, one needs a graph of Fig. 4 shown in your paper. Can you advise me as to what formula was used to construct the graph.*

Y.M. CHIN, U.K. : *The construction of Fig. 4 may be summarised by the following logical steps:*

(i) *Assume a value of holding time h.*

(ii) *Choose a value of traffic A for graph construction.*

(iii) *For the practical range of Cost Factor H (0 to 1), find the corresponding optimum high usage route N.*

(iv) *For s=o to A/10 (to preserve the inequality S/A≤1/10) find the corresponding penalty P=Δu/H and t_m, the measurement study period.*

(v) *Repeat 3 to 4 to obtain a data set for P and t_m, and find the B constant in the equation $P=1/Bt_m$.*

(vi) *Repeat 3 to 5 to obtain a B and H data set. A plot of B versus H will reveal a series of straight lines (see fig. 3 for example). The gradient, μ, of these lines can then be plotted for μ versus N.*

BIOGRAPHY

MUN CHIN studied at the Brighton Polytechnic, U.K., and was awarded an Honours Degree in Electrical Engineering in 1970. Subsequently he completed a Post-Graduate course at the Birmingham University, U.K., and gained a Master of Science Degree in Information and Systems Engineering in 1971. In the following year he was recruited by the P.M.G. from London, and was appointed to Traffic Engineering Section, N.S.W. His experience in the Planning Branch has been gained in the P.A.B.X. and Data Production Sub-Sections, where his work included teletraffic studies, design of the N.S.W. P.A.B.X. Data base, problems in teletraffic measurements, and forecasting. He is currently attached to Switching and Facilities, (METROP) N.S.W.

Applications of Processing State Transition Diagrams to Traffic Engineering

P. GERRAND

Telecom Australia*, Melbourne, Australia

ABSTRACT

The likely adoption by the VIth Plenary Assembly in October 1976 of the graphical Specification and Description Language (SDL) prepared by the CCITT's Study Group XI offers potential advantages to teletraffic engineers. This paper introduces the SDL, and suggests a general scheme for the systematic application of the SDL to system documentation, whereby the documentation required for capacity studies can be generated as a natural part of the system design process. The usefulness of processing state transition diagrams in general, of which the SDL is a special but important case, to both simulation and analysis of traffic capacity is discussed.

1. INTRODUCTION

Telephone switching systems have always been complicated things, and their increasing versatility with stored program control (SPC) has not made their behaviour any the easier to understand. One can highlight three stages in the life of an SPC telephone exchange in which deep understanding of its working is required. During the stages of acceptance testing weaknesses in the manufacturer's design or documentation, and weaknesses in the administration's specification, are usually discovered and corrected. During any major extension of the exchange to incorporate new facilities or attain a larger capacity, the system documentation may be found to be incomplete or inaccurate. Thirdly, whether carried out during the initial phase of system development or at any later period in the life of the exchange, a traffic capacity study using a simulation model or other means will only be accurate if the traffic engineers have absorbed a deep and detailed understanding of the system design. In all three stages, the accuracy, completeness and intelligibility of the system documentation are obviously of critical importance.

In February 1976, the CCITT's Study Group XI finalized a draft Recommendation on the use of a graphical Specification and Description Language (SDL) for application to the documentation of SPC switching systems. By the time this International Teletraffic Congress commences, it will be known whether the Plenary Assembly of the CCITT has approved this Recommendation or not. All the auguries point to its approval, in which case a new international standard will have been created, and one can expect that an increasing number of SPC switching systems will be documented using this SDL. Section 3 of this paper will provide an introduction to the SDL, and Sections 4 to 6 will indicate its potential usefulness to traffic engineers in the simulation, analysis and general understanding of switching systems - both SPC and non-SPC.

But first it will be necessary to devote a short section, Section 2, to a simple classification scheme for levels of system documentation, which will make it easier to explain the various applications of the SDL, and in particular to make clear what is meant by "specification" and "description".

---

* This paper was written, and is based upon work performed, while Mr. Gerrand was on leave from Telecom Australia, working as a consultant to the Laboratorios ITT de Standard Eléctrica, S.A. in Madrid, from November 1974 to October 1976.

2. LEVELS AND PARTITIONS IN SYSTEM DOCUMENTATION

In the context of the purchase of telephone switching systems, a clear distinction exists between a specification and a description of the behaviour of a system. A functional specification shows the requirements of a system in terms of its behaviour in response to several given sequences of inputs. A functional description shows the actual behaviour of a realized system in response to all possible sequences of inputs, giving considerable detail concerning the internal structure of the system.

Thus a specification normally precedes design, and views the behaviour of the system "from the outside", whereas a description normally follows design, and describes the behaviour of the system "from the inside".

However, in the context of the development of the system, the design process may be observed to pass through several phases, in which the same document may serve both as a description (of the design decisions made up to this phase) and as a specification (for further, more detailed design).

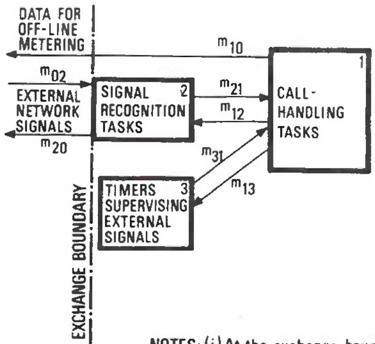
Amongst the many levels of documentation which characterise different phases of switching systems design, three levels appear to be fundamental, irrespective of whether the documentation language is of narrative or diagrammatic form.

A Level 1 document is a specification of the system requirements without prejudice to the choice of switching structure, control structure or technology. Level 1 is generally dependent only upon the interworking requirements between this and other parts of the network. Hence the CCITT R2 and No. 5 signalling specifications are examples of Level 1 documents, largely in narrative form.

A Level 2 document also specifies system requirements, but it is system-dependent to the extent that it introduces the design decisions concerning the structure of the switchblock (e.g. junctors, switching units and signalling units) and the peripheral units (e.g. markers, scanners and drivers) that will drive the switchblock. Level 2 does not introduce any design decisions concerning the internal structure of the Central Control. Hence Level 2 serves as a description of the system behaviour from the point of view of its switching structure, and as a specification of the behaviour of the Central Control.

Level 3 introduces the design decisions concerning the internal structure of the Central Control (e.g. the internal queueing structure, and the allocation of central memory blocks to call processes, etc.), as well as all design decisions described in Level 2. Thus Level 3 serves as a high-level description of the behaviour of processes in the system; "high-level" by contrast with the flowcharts and lists of coding that serve as the implementation levels in design.

Within any of these Levels, it is convenient to partition the system documentation into several Functional Blocks (using the terminology of the CCITT SDL). Examples of the typical Functional Blocks one might expect to find in each of the three Levels are shown in Figs. 1, 2 and 3. Each Functional Block has only a partial view of what is occurring in the whole system, and each Functional Block must maintain its own consistent point of view: e.g. per-call, per-equipment or per-subsystem. Through the interaction of the Functional Blocks, the behaviour of the total exchange is specified or described.



NOTES: (i) At the exchange boundary, data formats for the signals must be specified, but not at the internal boundaries - to avoid prejudicing the design.
(ii) m_{ij} represents the set of signals required from Functional Block i to Functional Block j.

Fig. 1 Example of the partitioning of a Level 1 switching system specification into Functional Blocks

Further we note that a full description of the system in Level 3 will normally include Functional Blocks that are never specified in Level 2, but are left to the choice of the designers of the Central Control, e.g. the internal queueing structure of the system. Similarly, the system designers will normally specify certain Functional Blocks in Level 2 that were never specified by the customer in Level 1, e.g. the system recovery and initialization tasks.

The structure of Levels and Functional Blocks which has been introduced in this Section is independent of the language chosen for specification and description; it serves as a framework in which the techniques of a graphical language such as the CCITT SDL can be flexibly employed.

3. INTRODUCTION TO THE SPECIFICATION AND DESCRIPTION LANGUAGE

3.1 THE CONCEPT OF A PROCESS

The central concept in the CCITT SDL is that of a process. This word is not formally defined; its meaning is entirely compatible with its use in ordinary English, and also with its use in the theory of stochastic processes. Some relevant examples of processes are: any kind of telephone call process; a signal-recognition process; a system recovery process. In short, the names of the different Functional Blocks shown in Figs. 1, 2 and 3 indicate the nature of the processes belonging to these Blocks. A Functional Block may contain one or more processes, but to avoid confusion we insist that any given process may only belong to one Functional Block, in any given Level.

It is typical of a process that it passes through several states. Often it may be considered as a life-and-death process, in which its birth occurs when it leaves some "idle" or "initial" state, and its death occurs when it returns to that original state or goes to some permanent terminal state; such a process may have any number of lives. For example, a call-process may be created by assigning a word of memory to an incoming line, initialised with a memory state meaning "line equipped, in service, and idle". Each time the memory state changes from "idle" to "seizure", a new call is said to be born and only when it returns to the idle state is that call considered to have died.

3.2 BASIC DEFINITIONS IN THE CCITT SDL (Ref. 1)

Signals: a signal is a flow of data conveying information to or from a process. A signal may be either in hardware or in software form. If the information flow is from a process described by a Block to a process described by another Block it is an external signal. If the flow is between processes described by the same Block it is an internal signal.

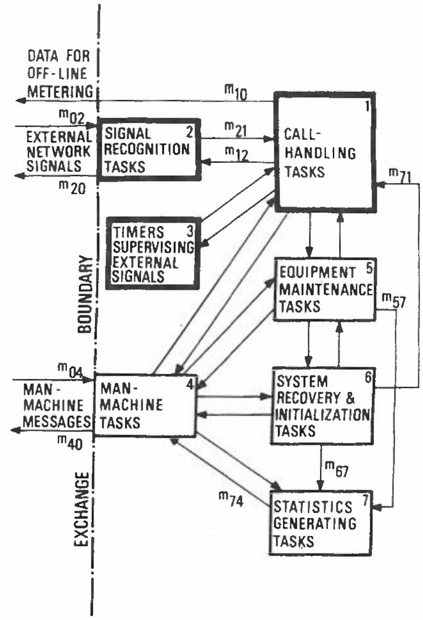


Fig. 2 Example of the partitioning of a Level 2 switching system specification into Functional Blocks (Compare with Fig. 1)

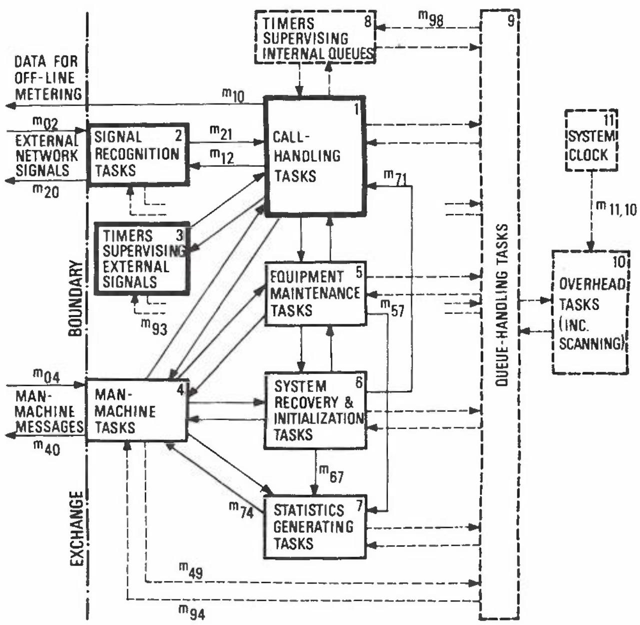


Fig. 3 Example of the partitioning of a Level 3 switching system description into Functional Blocks

Inputs: an input is an incoming signal which is recognized by a process. In accordance with the definition of signals, an input can be internal or external.

States: a state is a condition in which the action of a process is suspended awaiting an input.

Transitions: a transition is a sequence of actions which occurs when a process changes from one state to another state in response to an input. (Hence at any given instant, a process can only be either in one of its states or else in a transition). Note that "transition" has been given a processing-oriented meaning that does not exist in the simple state-transition diagrams well-known to teletraffic engineers in application to birth-and-death equations.

Outputs: an output is an action within a transition which generates a signal which in turn acts as an input elsewhere. In accordance with the definition of signals, an output can be either internal or external; but note that the output is the processing action which generates the signal, and not the signal itself.

Decisions: a decision is an action within a transition which asks a question to which the answer can be obtained at that instant and chooses one of several paths to continue the transition.

Tasks: a task is any action within a transition which is neither a decision nor an output.

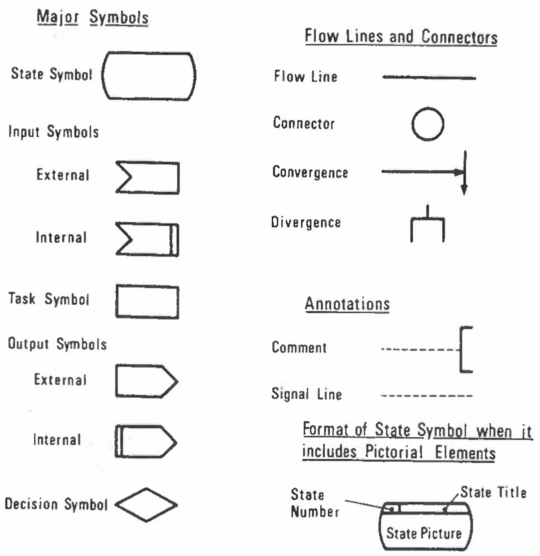


Fig.4 Recommended Symbols in the CCITT SDL

3.3 SYMBOLS OF THE CCITT SDL

The basic and recommended symbols of the CCITT SDL are shown in Fig. 4. These graphical symbols are combined to create a diagram that will describe the behaviour of a process. Such a diagram may be called a processing state transition diagram (PSTD): a PSTD is a graph which describes the (required or observed) behaviour of a process in terms of its set of possible states, the processing transitions between these states, and the inputs which cause these transitions. The CCITT SDL is but one of several graphical languages that have been proposed for PSTDs, but it has the merit of being the only language to have achieved international standardization. Particular cases of a PSTD, when applied to a call, an equipment, a queue or to overhead (inter-queue) logic, will be referred to as a CSTD, ESTD, QSTD and OSTD respectively, irrespective of the particular graphical language used to represent the processing transitions*.

To be a legitimate PSTD using the CCITT SDL, a necessary condition is that each transition must satisfy the connectivity diagram shown in Fig. 5, starting and ending at a state symbol, passing only once through an input symbol. This condition eliminates illegal combinations such as a dangling transition that never terminates in a state, or transitions without inputs, etc.

The CCITT SDL does not restrict the contents of its major symbols; it specifically includes the option of including state pictures within a state symbol as a means of identifying the exact state of a process and also of simplifying the definition of a sequence of processing actions within a transition. An example of a Level 1 specifica-

* The Japanese use the term STD in a sense sufficiently general to include all PSTDs.

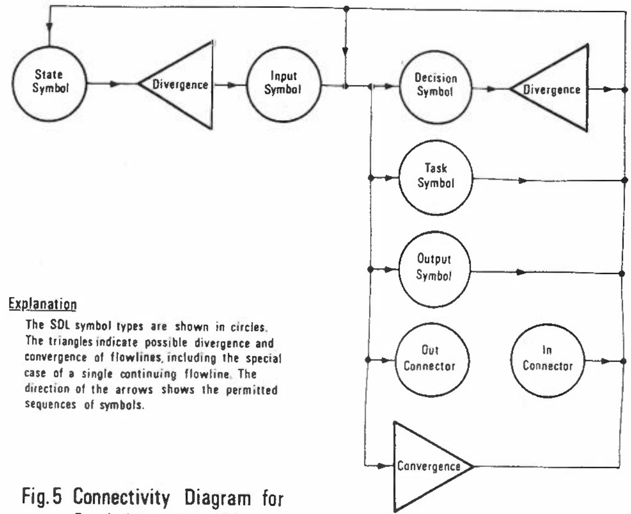


Fig.5 Connectivity Diagram for Symbols in the SDL

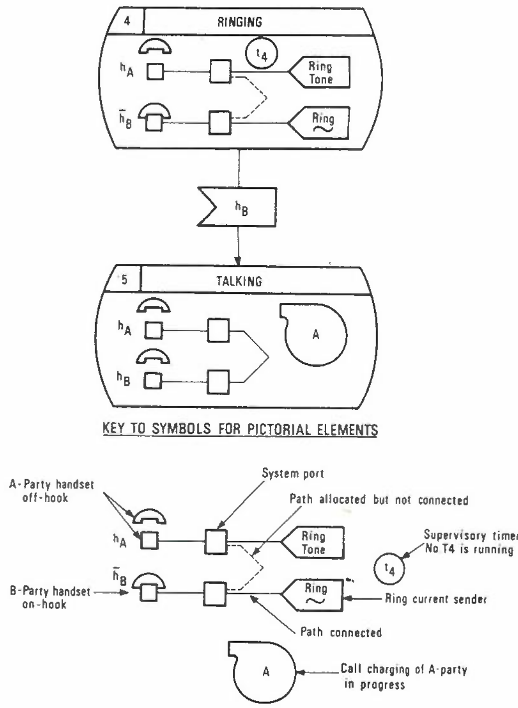


Fig.6a Typical use of state pictures in the specification of call-handling processing. (Level 1)

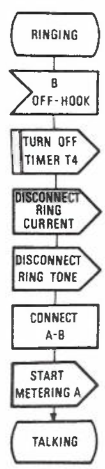


Fig. 6b Specification (Level 1) equivalent to Fig.6a, using the SDL without state pictures.

tion of a call-processing transition is shown in Fig. 6. The processing required in going from state 4 (ringing) to state 5 (Talking) is the same in Fig. 6a, using state pictures, as in Fig. 6b, not using state pictures. As shown in this example, the total processing involved when passing from one state to the following state is that required to effect the changes in the state pictures, together with the processing indicated in any decisions, outputs or tasks appearing in the transition between the states.

The CCITT has not yet recommended the use of particular symbols for the pictorial elements appearing within state pictures: this is the subject of further study by the CCITT's Study Group XI in 1977-80. This paper uses the symbols for pictorial elements that have been preferred in trial use within the Australian administration (Refs. 2-4); for alternative symbols, see Refs. 5-7.

4. SYSTEMATIC APPLICATION OF THE SDL TO SYSTEM DOCUMENTATION

The greater usefulness of the SDL to teletraffic engineering is this: that the documentation required for simulation studies and system analysis can be generated as a natural part of the system design process, with equal value to system designers and traffic engineers.

Fig. 7 proposes a scheme for the systematic application of the SDL to system documentation. It assumes a situation in which the customer has no direct involvement with the design of the switching system, and wishes to provide a system independent specification*. Functional system specification begins with a set of Level 1 PSTDs, negotiated between customer and supplier. It is expected that all the CCITT international signalling specifications will appear in SDL form in the next few years, and that national signalling specificatons will be similarly expressed. To show the power of expression of the SDL, a Level 1 specification for R2-R2 transit signalling will be provided as a handout (Appendix 1) to this paper. For examples of the use of XSTDs in the Level 1 specification of sophisticated local call facilities, see Gale (Ref.4).

Within the manufacturer's organization, a system specification group can prepare a set of Level 2 PSTDs in which the hardware structure appropriate to a chosen generic design is incorporated. This set of documents needs to be negotiated between the system specifiers and designers before reaching final status as the Level 1 specification for the designers responsible for the Central Control.

The Level 2 CSTD specifications can be produced by a process of systematic modification of the Level 1 CSTDs, requiring relatively little effort. The modification consists of four steps (see Fig. 8 parts (a) and (b)):

(a) Modify the Level 1 state pictures to reflect the choices of switching structure and signalling equipment modules;

(b) In the Level 1 transitions, insert the appropriate orders for the markers, drivers and other equipment peripheral to the Central Control**;

(c) Relate the signalling states in the Level 1 CSTDs to hardware testpoint conditions in the chosen equipment; then augment the CSTD to include transitions due to those extra testpoint conditions (usually fault conditions) not considered in Level 1;

---

* When the customer is directly involved in the design of the switching system, as occurred in the development of the Japanese D10 system, it can be more convenient to begin functional specification with an agreed switching structure (trunking diagram), and hence commence with Level 2 PSTDs. Several examples of such Level 2 specifications are given in Refs. 5-7.

** This step is not strictly necessary if the changes in the state pictures alone are sufficient to determine unambiguously the identities of the peripheral equipment needed to effect these changes as well as the appropriate orders.

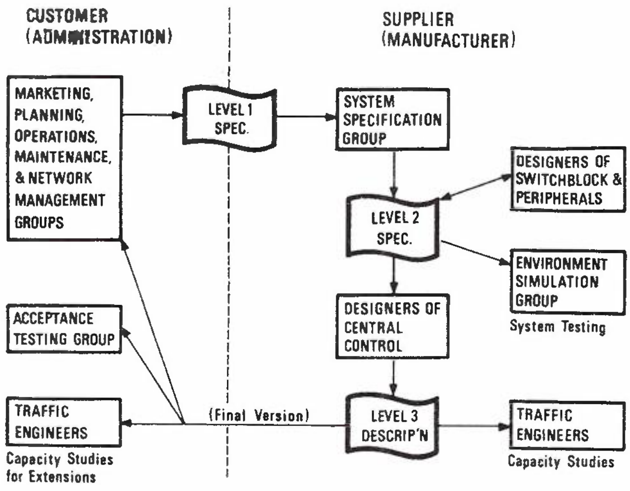


Fig.7 Flow of Level 1, 2 and 3 documents using PSTDs (Processing State Transition Diagrams)

(d) Augment the CSTD to include those equipment-testing routines (such as path continuity tests) that the system designers decide to include as an intrinsic part of a call process.

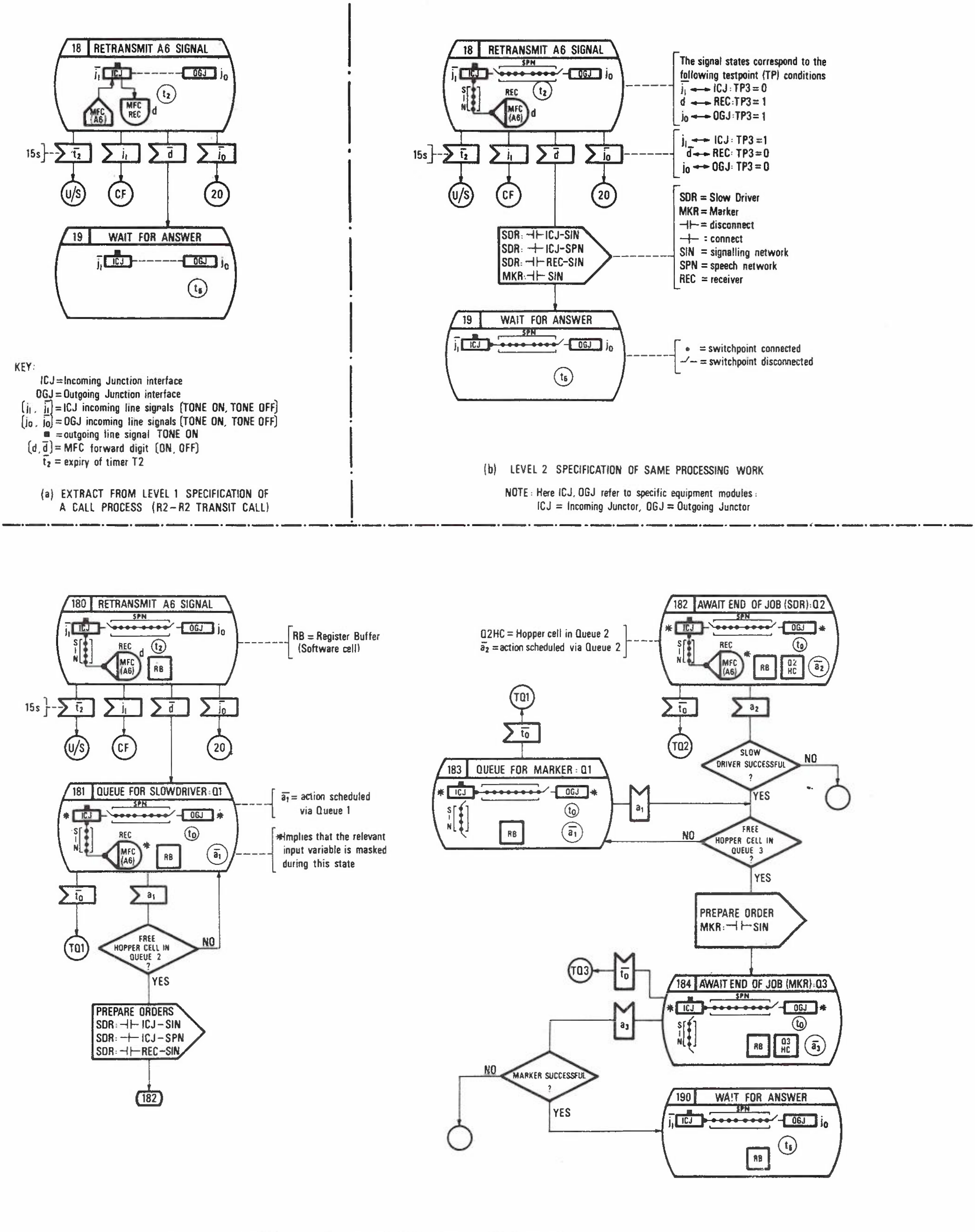
Once completed, the Level 2 PSTDs serve not only as a specification for the design of the Central Control, but also as a specification for the environmental simulation team, whose job is to test the functioning of the Central Control by simulating its exchange environment. This is made possible because the Level 2 PSTDs include not only the external exchange signals (as in Level 1), but also the output orders to the equipment modules peripheral to the Central Control (see Fig. 8b). In addition, the Level 2 CSTDs can be annotated by traffic engineers to show both call-mix data (the probability that the call will follow any particular transition from a given state) and call-pattern data (the average arrival time of each input signal, plus any additional parameters used to model its arrival time distribution),as valuable input data for the environmental simulation.

The Level 2 PSTDs do not suffice as a system description for capacity studies, since they do not describe the queueing logic that services calls within the system, nor the states in which a call is queued for further processing work, nor the fates of calls when they recognize inputs such as time-outs during queueing states.

This information is found in Level 3 PSTDs, which describes the complete behaviour of processes in the realized system, including all fault conditions except those unpredictable processor faults which would invalidate the description itself. Ideally, from the point of view of traffic engineers, Level 3 PSTDs can be produced by the designers of the Central Control themselves as a natural part of the design process, preceding the design of program flowcharts and coding. Once customers start demanding a full system description using the CCITT SDL, it is likely that the designers will be given the responsibility for the production of the Level 3 diagrams. If the designers are committed to a design process using other documentation techniques, the Level 3 PSTDs can be produced retrospectively as a higher-level description of the logic which the designers have already implemented. This may seem a formidable task, but it can be simplified enormously if the starting point is the Level 2 documentation.

To convert Level 2 PSTDs to Level 3 PSTDs, the following steps can be performed systematically (see Figs. 8b & 8c):

(a) Extend the Level 2 state pictures to include those software cells and timers that are allocated to the process during these states. (If a software cell is allocated permanently to a process, e.g. a "line



(a) EXTRACT FROM LEVEL 1 SPECIFICATION OF A CALL PROCESS (R2-R2 TRANSIT CALL)

(b) LEVEL 2 SPECIFICATION OF SAME PROCESSING WORK

NOTE: Here ICJ, OGJ refer to specific equipment modules: ICJ = Incoming Junctor, OGJ = Outgoing Junctor

(c) LEVEL 3 DESCRIPTION OF THE PROCESSING WORK SPECIFIED IN (a) AND (b)

FIG. 8 COMPARISON OF LEVEL 1, 2 AND 3 DOCUMENTATION OF THE SAME PROCESSING WORK

equipment category word" for a call, there is no point in adding this information to the state pictures).

(b) Draw and define extra states corresponding to all suspended processing states that may occur in between the Level 2 states. Their state pictures should indicate that further processing action has been scheduled.

(c) For each state in the PSTD, identify the effects of all possible inputs (particularly line signals, registrer signals, time-outs and further processing actions), showing the processing transitions and the new states reached in every case. The processing transitions should include all decisions that affect the fate of a call, e.g., that decide the extent to which it will be delayed, or whether it will succeed or fail.

In practice, it is useful to annotate the flowlines in the Level 3 PSTDs so as to identify the software programs which are utilized in each transition. In this way, processing actions summarized concisely in the Level 3 PSTDs can be checked against their more detailed descriptions in the lower level system documentation. (These annotations are not shown in Fig. 8c).

5. APPLICATION TO SYSTEM SIMULATION

Perhaps the most general and basic difficulty encountered in the simulation of a switching system occurs when the implementation documentation is not suitable as a functional description for simulation modelling. The traffic engineer is confronted with the flowcharts for perhaps 100 k words of program code; he has to separate the relevant from the irrelevant, putting aside those programs and routines which will not affect system capacity; he has to re-orient the system description from one concerned with the efficient execution of modularized switching functions, to one primarily concerned with the frequencies of delay and loss of calls through the system. If he does not adequately re-orient the system description, he will become overly preoccupied with measurements that are meaningful to the system designers, but which have only secondary relevance or even no relevance whatsoever, to the calculation of system capacity.

Then, when the traffic engineer has designed and coded a compact simulation model of perhaps 5 k statements, it is likely that his own system description - of the simulation model - will be heavily oriented to the particular computer simulation language used, and hence not readily comprehended by the system designers, who will have difficulty in checking the simplifying assumptions introduced into the simulation model, if they are indeed brave enough to attempt to check them.

This may seem a gloomy scenario, but I believe it is more typical than untypical of the state of the art of switching system simulation in the world today: in which the traffic engineer has great difficulty in penetrating the documentation of the real system, and in which the system designers have equal difficulty in penetrating the documentation of the traffic simulation model. This difficulty in communication does not prevent reasonably accurate estimates being obtained for system capacity, but is does lead to the following typical costs and overheads:

(a) changes in the system design usually necessitate changes in the coding of the simulation model, rather than simply changes in the input data to that model;

(b) effectively a new simulation model is designed for each significantly different network application of a switching system.

The thought occurs that, if it is reasonable to assume that the behaviour of any contemporary SPC switching system can be described using a set of PSTDs, then a machine which can read a set of PSTDs and then simulate the behaviour of that set of PSTDs will be a truly generic simulation model, capable of simulating any contemporary SPC switching system. This machine would not only be free from the costs and overheads mentioned in the previous paragraph; it would also have the advantage that the PSTDs can serve as common documentation for the real system and the simulation model itself.

The appropriate documentation would consist of Level 3 PSTDs in which estimates of processing times are given for each significant transition. The significant transitions are those whose probabilitites of execution, from the point of view of traffic engineering, are not negligible, and whose execution affects the traffic capacity of the exchange. Exclusion of non-significant transitions from consideration is motivated as much by the desire to minimize the work of the system designers in estimating processing times as to reduce the quantity of input data to the simulation model.

Since the CCITT SDL is currently applicable to sequentially functioning processes only, it would be necessary to first partition the system into its minimum number of sequential machines: thus each central processor, each peripheral device, and even the autonomous system clock may need to be identified separately. Having made this first "horizontal" partition of the system, a second "vertical" partition of the system would be advantageous (see Fig. 9), dividing each sequential machine into three Functional Blocks, whose Level 3 PSTDs are called:

(i) the OSTD (Overhead STD), which describes the inter-queue logic, including for example the scanning logic in a central processor;

(ii) the QSTD (Queue STD), which describes the intra-queue (queue-handling) logic for each queue in this sequential machine;

(iii) The CSTD (Call STD), which describes the call processes as handled by this sequential machine.

The partition of the total queueing structure into an OSTD and QSTD is particularly convenient for the central processor, in which a multi-queue structure typically occurs. (In the case of the peripheral devices and the system clock, no OSTD is necessary if each of these devices acts as a single-queue or zero-queue subsystem: see Fig. 9). It is believed typical of SPC system design that individual changes in the system design will tend to create changes in only one of the OSTD, QSTD or CSTD but rarely two of these at a time. For example, a change in the overload control strategy is likely to necessitate a change in the OSTD for the central processors alone; and the addition of a new call-handling facility is likely to change the CSTDs alone.

An example of the interaction between parts of the OSTD, QSTD and CSTD for a central processor is shown in Fig. 10.

In order to prepare the diagrams as input data to the simulation, it is of course necessary to obtain estimates of the processing times in all significant transitions in the Level 3 PSTDs. In practice, most of these estimates will be constant times; a small minority will require modelling with a particular random distribution of processing times.

Since the purpose of this Section is to be suggestive rather than prescriptive, it is not intended to go into details of the preparation of the Level 3 PSTDs as input data. Suffice to say that each significant transition in a Level 3 PSTD can be assigned an input data "cell", consisting of a well-defined list of parameters, such as: transition time; output signals to be sent to other processes; the identities of queues or devices whose occupancy counters must be incremented or decremented during this transition; the next state of the process; and the identity of the PSTD to which control is returned. In cases where any of these parameters are not predetermined, special characters can be used to identify library subroutines which will generate the required data.

During the author's work as a consultant engineer in Madrid in 1975-76, he has contributed to the design of a generic simulation model with a more modest range of application, being content to simulate the behaviour of a particular SPC transit exchange in a wide range of network applications. The scheme of partitions shown in Fig.9

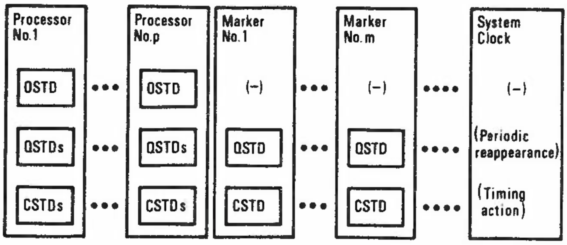


Fig.9 PSTDs for the Simulation Model: Partition of the exchange "horizontally" into sequential machines and "vertically" into Functional Blocks

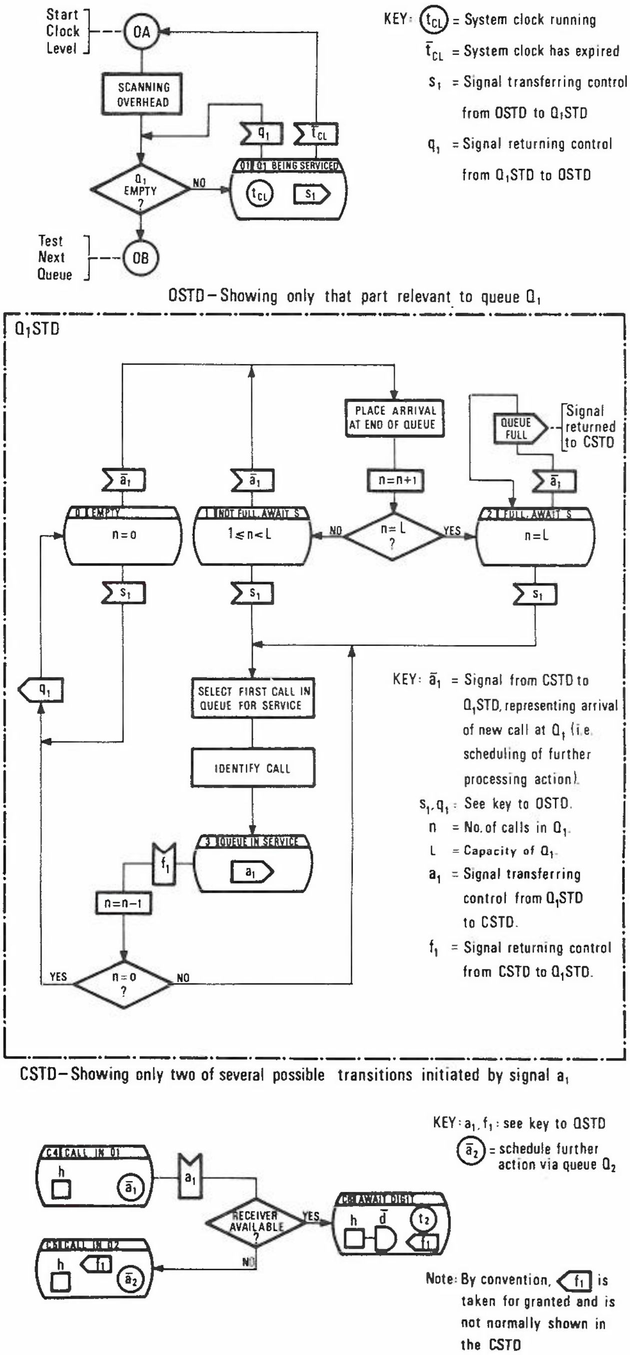


Fig.10 Interactions between OSTD, QSTD and CSTD in a Central Control processor

was used, and the starting point for the development of the Level 3 PSTDs was the set of Level 2 CSTDs produced by the system specification group. The traffic engineering team found it possible to develop a complete set of relevant CSTDs, QSTDs and OSTD many months in advance of the coding of the real system, by consultation with the system designers. Preliminary estimates of processing times have been inserted in the input data to the simulation model, in advance of the accurate estimates obtainable when the system is fully implemented. In this way, the traffic engineering team is able to produce results which will influence the detailed design decisisons, before the system has been implemented as a "fait accompli".

6. APPLICATION TO ANALYSIS

Since the previous International Teletraffic Congress in 1973, several papers have been published, notably Refs. 9 and 10, on the difficult problem of analysing the traffic capacity of an entire local telephone system or telephone network, and its performance during overload conditions. Since these global analyses must take into account customer behaviour, traffic dispersion and all significant forms of congestion in the network, the algebraic forms of their solutions are inevitably very complicated.

It is the author's belief that such global analyses can be significantly aided if a Level 3 CSTD is used as the basic model to define the range of dynamic behaviour of the call processes in the given system. The Level 3 CSTD has the following relevant properties:

(a) it shows every possible fate of the call (excluding those extreme processor fault conditions which would violate the logic of the CSTD) and hence includes the effects of all relevant call-failure mechanisms;

(b) its state pictures enable one to relate the analytical model to the allocation of equipment, supervisory timers and queues to calls in the real system, and hence to manipulate the analytical solution according to changes in the system design;

(c) its properties as a directed graph (or signal-flow-graph) can be exploited to produce a formula for the probability of call-failure, taking into account all significant mechanisms of call-failure.

These ideas have been explored in an earlier paper by the author (Ref. 3); a brief outline will be given here to show their general thrust.

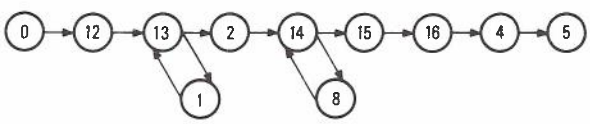


Fig. 11 Signal-flow graph showing all call set-up paths in a given Level 3 CSTD (Ref.3, Fig.1)

A call's progress can be viewed as a random path through a CSTD, starting at the idle state and passing from state to state via the permitted transitions. As an example, Fig. 11 is a subgraph of a Level 3 CSTD (Ref. 3, Fig. 1), reduced to nodes and flowlines, showing the permitted paths from the idle state (node 0) to the conversation state (node 5). Any CSTD using the CCITT SDL can be reduced to such a simple signal-flow graph by replacing the state symbols and decision symbols with numbered nodes, absorbing the other major SDL symbols (inputs, outputs and tasks) into the flowlines that connect them, and discarding all flowline and nodes that are not part of call set-up sequences.

A probability function $P_{m,n}$ can be associated with each flowline in a CSTD's signal-flow graph:

Definition 1: $P_{m,n}$ = probability that a call currently in node m will make its next transition to node n.

Using the convenient algebra of signal-flow graphs, the total probability of reaching node 5 from node 0 in Fig. 11 can be written down by inspection:

$$P^T_{0,5} = \frac{P_{0,12}P_{12,13}P_{13,2}P_{2,14}P_{14,15}P_{15,16}P_{16,4}P_{4,5}}{(1-P_{13,1}P_{1,13})(1-P_{14,8}P_{8,14})} \quad (1)$$

The system's probability of loss for this type of call might typically be defined as $(1-P^T_{0,5})$ under the assumptions that the calling subscriber is "in service" ($P_{12,13}=1$) and actually attempts a call ($P_{0,12}=1$). The problem remains of relating the individual transition probabilities $P_{m,n}$ to the basic parameters of the telephone network.

In the cases of signal flow from a decision node, the transition probabilities can generally be evaluated either from traffic dispersal statistics (for such decisions as DIGIT ANALYSIS) or by applying classic traffic theory to link congestion in the exchange (for such decisions as DIGIT RECEIVER AVAILABLE?).

In the case of signal flow from a state node, the transitions are triggered by inputs, which are time-dependent events; the probability of a call taking a particular transition corresponding to event #1 is of course the probability that event #1 occurs before any of the events #2, #3, ...#N.

Definition 2: P{#1<(#2,#3,...#N)}= the probability that event #1 occurs before any of the events #2, #3.. and #N.

Definition 3: $q_i(t)$= the probability that event #i occurs in the time interval (0,t).

Theorem 1:

For N statistically independent events #1, #2,...#N,

$$P\{\#1<(\#2,\#3,\dots,\#N)\} = \int_0^\infty \dot{q}_1 \prod_{i=2}^{N} (1-q_i)\,dt \quad (2)$$

where the time origin is chosen so that $q_1(0)=0$ for all i.

Ref. 3 proves this simple theorem and gives tables of results for (a) pairs of competing events, and (b) trios of competing events, when several arrival distributions, typical for telphony events, are substituted for the $q_i(t)$.

NOTE: As a consequence of Theorem 1, an additional result, useful for calculating device occupancy, cna be given in terms of the already defined probability functions. For a given state with N statistically independent input events, the time τ spent in that state can be given a cumulative distribution function $F(t)=P\{\tau<t\}$, evaluated as

$$F(t) = \sum_{i=1}^{N} q_i(t)\cdot P\{\#i<(\text{all}\#k{:}k\neq i)\}$$

$$= \sum_{i=1}^{N} q_i(t)\cdot \int_0^\infty \dot{q}_i(t)\cdot \prod_{k\neq i}(1-q_k(t))\,dt \quad (3)$$

Since each state picture indicates the devices occupied during that call state, equation (3) can be used to calculate the distribution function for device occupancy times, if it is valid to assume that the time the call spends in transitions is negligible compared to the time it spends in states.

7. CONCLUDING REMARKS

This paper has introduced a general classification of three levels of system documentation in order to aid discussion of the applications of the CCITT SDL to traffic engineering. (It should be noted that the CCITT's draft Recommendation, Ref. 1 does not itself include any classification scheme for levels of system documentation, although it assumes that the SDL will be applied to more than one level within a hierarchical scheme of system documentation).

Perhaps the key suggestion in this paper is the systematic application of the CCITT SDL to the specification, design and description of processes in SPC switching systems, using the scheme illustrated by Fig. 7. This scheme enables the documentation required by traffic engineers for capacity studies to be generated as an integral part of the design process. The same documentation, in its final version, also serves the customer as a high-level description of the behaviour of the system. This proposal is put forward with the hope of benefiting both the customer and the supplier; it is the suggestion of the author as an individual, and has yet to be considered within Telecom Australia.

8. ACKNOWLEDGEMENTS

The CCITT SDL, Ref. 1 was prepared by Sub-Group XI/3-1 of Study Group XI in 1973-76, with the active participation of experts from more than fifteen telecommunications administrations and manufactureres. The author is grateful to Standard Eléctrica, S.A. for releasing him at intervals from his work as a consultant in 1975-76 to attend meetings of Sub-Group XI/3-1 as a nominated expert of the Australian Administration. In developing ideas for the systematic application of the SDL to system documentation, the author has benefitted from several interchanges of ideas with other members of Sub-Group XI/3-1. Most of all, as in his previous work (Refs. 2 and 3), the author has been deeply influenced by the multiple possibilities inherent in the use of state pictures, which originated in the early stages of the Japanese DEX projects (Refs. 5 and 6).

For experience gained in applying PSTDs to the specification, description and simulation of an SPC international transit exchange, the author is grateful to his former colleagues in Standard Eléctrica: Dr. Clementina Bravo for her help in writing the Level 1 CSTDs; Brian Andrews and Luis García Fernández for their work in writing the Level 2 CSTDs; and Susi Sánchez-Puga, Enrique Martí and Dr. Clementina Bravo for their help in producing the Level 3 PSTDs.

Finally, for fortifying the author's motivation to persevere with the struggle to achieve a massive improvement in the quality of system documentation for the purpose of traffic engineering, he is grateful to Dr. Clem Pratt in Australia, and to Antonio Guerrero and Eduardo Villar in Spain.

9. REFERENCES

(1) CCITT, "Draft Recommendations Z.101-103", in the Third Part of the Final Report of Study Group XI to the VIth Plenary Assembly, Document AP6 No. 18, May 1976.

(2) The Australian Administration, "Call-State Transition Diagrams as a Specification Method for SPC Telephone Exchanges", CCITT Document Com. XI No. 26, Oct. 1973.

(3) P.H. Gerrand, "Use of Call-State Transition Diagrams in the Analysis of Telephone Call Failure Probabilities", Aust. Telecommn. Res., Vol. 8 No. 1, 1974, pp. 7-12.

(4) N. Gale, "Application of State Transmission Diagrams to System Independent Specification of Telephone Facilitites and Systems", Aust. Telecommun. Res., Vol. 9 No. 2, 1975, pp. 3-12.

(5) H.Kawashima, K. Futami and S. Kano, "Functional Specification of Call Processing by State Transition Diagrams", IEEE Trans., Vol. Com. 19 No. 5, Oct. 1971, pp. 581-7.

(6) NTT Japan, "Program Logic for Stored Program Controlled Telephone Exchanges", CCITT Document Com. XI No. 6, Dec. 1969.

(7) NTT Japan, "Functional Specification Descriptive Languages for Telephone Switching Systems (State Transition Diagrams)", CCITT Document Com. XI No. 46, March 1974.

(8) R.L. Franks and R.W. Rishel, "Overload Model of Telephone Network Operation", Bell Sys. Tech. J., Vol. 52, No. 9, Nov. 1973, pp. 1589-1615.

(9) E. Szybicki, "Determination of the Overload Capability of Local Telephone System", Electrical Communication Vol. 49 No. 2, 1974, pp. 170-176.

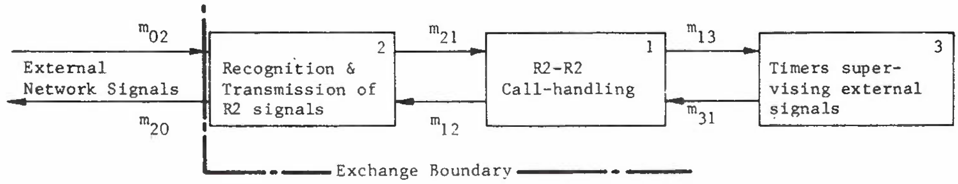
LEVEL 1 SPECIFICATION OF R2-R2 TRANSIT SIGNALLING (ANALOGUE VERSION) ITC8/Paper 313/APPENDIX I

(This specification comprises Appendix I to the paper "Applications of Processing State Transition Diagrams to Traffic Engineering" by P.H. Gerrand).

1. INTRODUCTION

This Appendix demonstrates the application of the CCITT SDL (Specification and Description Language) to the specification of the call-handling tasks required in R2-R2 transit signalling (with the analogue version of line signalling). The graphical specification of the call-handling tasks, shown in Section 4 below, has been extracted from the CCITT narrative Recommendations Q.350-Q.356 and Q.361-Q.368 of the Green Book, Vol. VI, 1972. The effect of pilot failures in the transmission equipment on call-handling has been omitted from this graphical specification, for convenience in preparing this example, but otherwise all sequences of line and register signals that impinge upon call-handling are believed to have been taken into account.

That part of the telephone exchange concerned with R2-R2 transit signalling is considered to be partitioned functionally into three Functional Blocks, as follows, without prejudice to methods of implementing the system.



2. INPUT AND OUTPUT SIGNALS RELEVANT TO FUNCTIONAL BLOCK 1

2.1 The input signals recognized by the call-handling processes are as follows :

(a) Set $m_{21} = \{ j_I, \bar{j}_I, j_O, \bar{j}_O, d, \bar{d}, b, \bar{b} \}$

where j_I = incoming junction (line) signal TONE ON; j_O = outgoing junction (line) signal TONE ON;
$\bar{j}_I$ = incoming junction (line) signal TONE OFF; $\bar{j}_O$ = outgoing junction (line) signal TONE OFF;
d = (forward) multifrequency code (register signal) DIGIT RECOGNIZED.
$\bar{d}$ = (forward) multifrequency code (register signal) NO DIGIT RECOGNIZED ("NO TONE").
b = backward multifrequency code (register signal) DIGIT RECOGNIZED.
$\bar{b}$ = backward multifrequency code (register signal) NO DIGIT RECOGNIZED ("NO TONE").

(b) Set $m_{31} = \{ \bar{t}_i \mid i = 1 \text{ to } 11 \}$

where $\bar{t}_i$ = expiry of the corresponding supervisory timer T_i.

2.2 The output signals from the call-handling processes are defined pictorially, as explained in the following Section 3.

3. KEY TO PICTORIAL ELEMENTS IN THE STATE PICTURES

ICJ = interface with incoming junction (ICJ), with backward line signal TONE ON.

ICJ = interface with ICJ, with backward line signal TONE OFF.

OGJ = interface with outgoing junction (OGJ), with forward line signal TONE ON.

OGJ = interface with OGJ, with forward line signal TONE OFF.

MFC (A6) = multifrequency code (MFC) sender, transmitting register signal A6.
(Other abbreviations for MFC signals include: NT= No Tone; INFO TONE= information tone, such as busy tone).

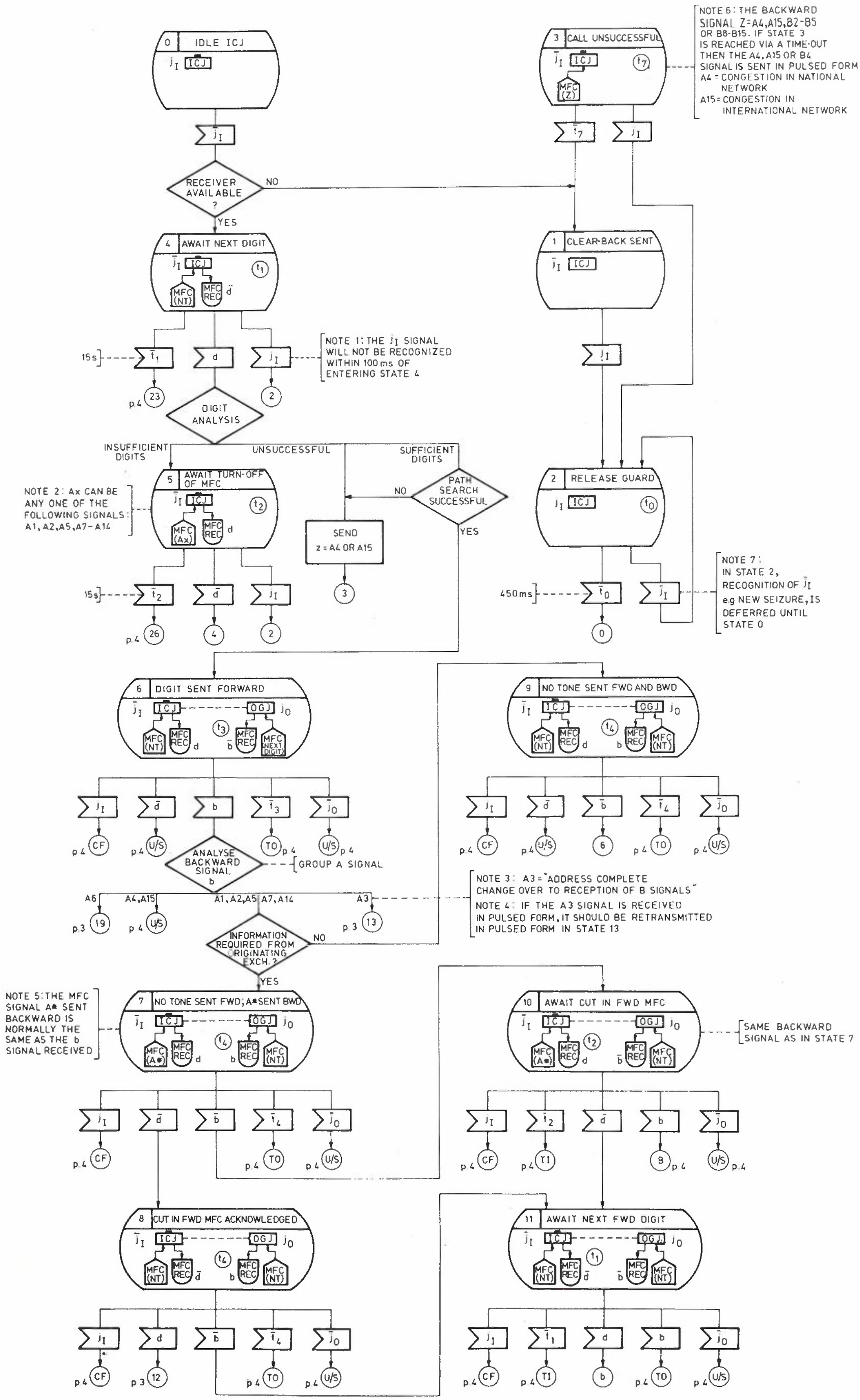
NOTE : Tables 2, 3, 4 and 5 of CCITT Rec. Q. 361 should be consulted for the explanation of the meaning of each register signal.

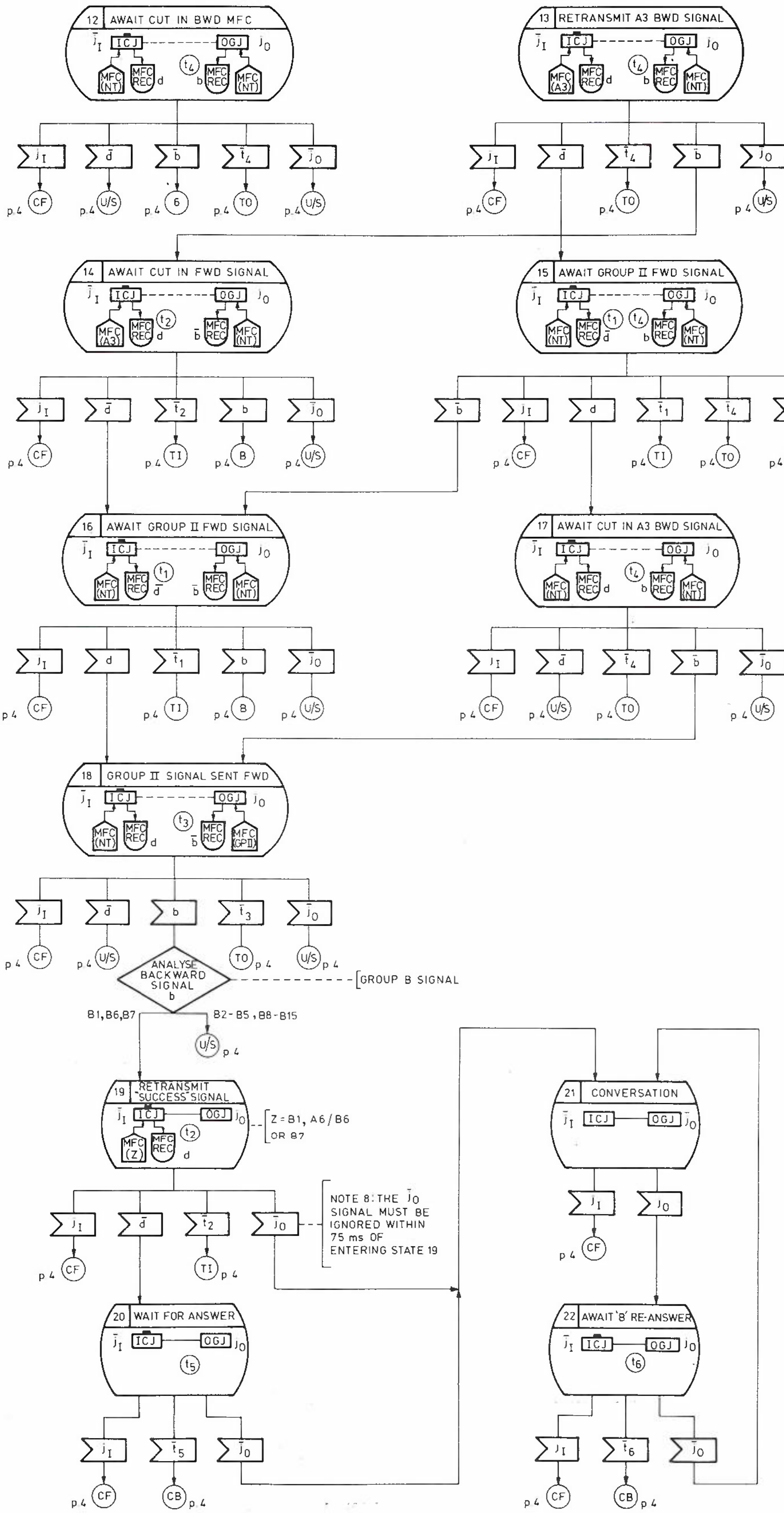
MFC REC = MFC receiver. (For signal conditions d, $\bar{d}$, b and $\bar{b}$, see section 2.1 above)

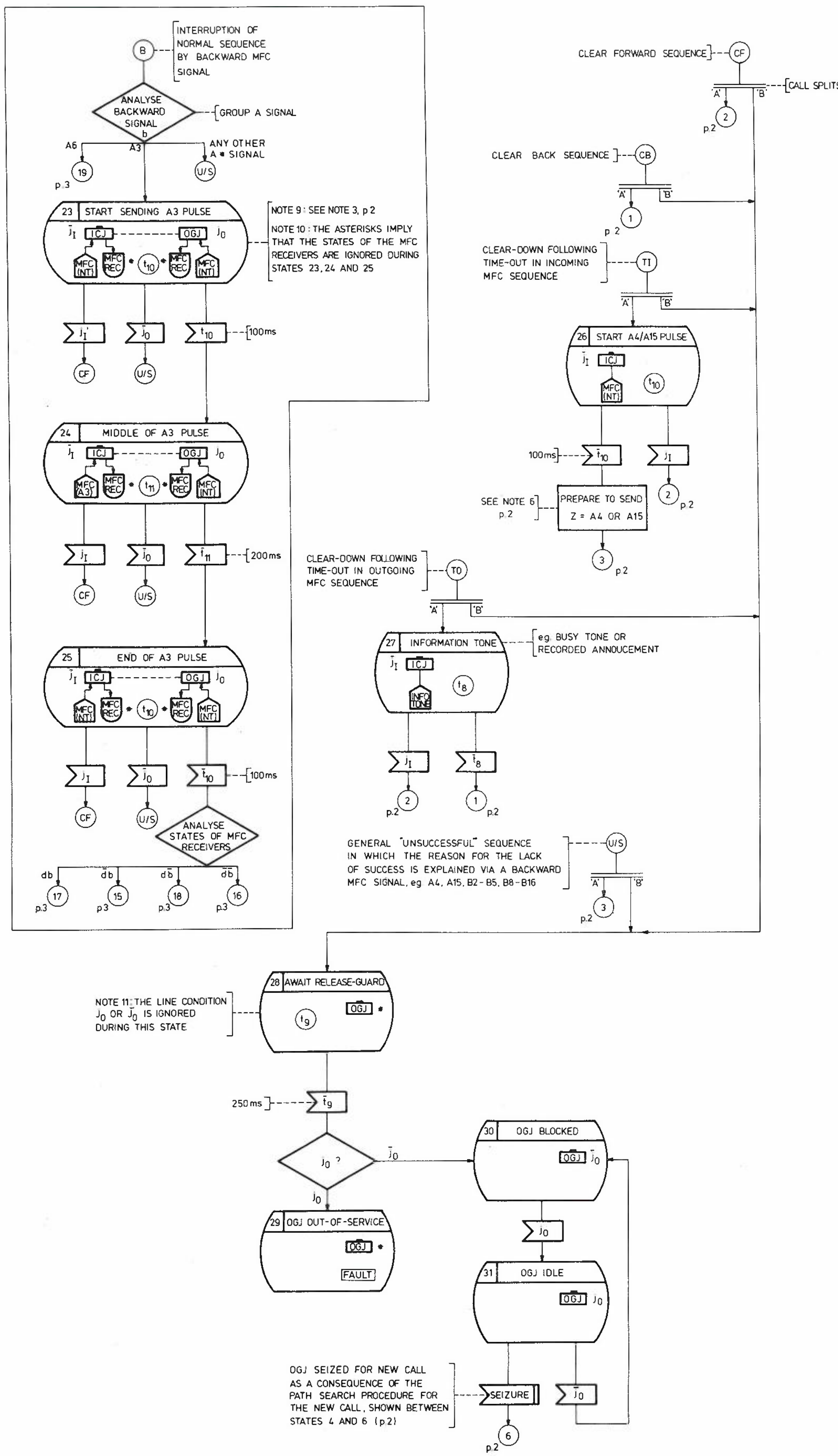
——— = path connected, whereas ------------ = path reserved but not connected.

t_i = supervisory timer T_i is running.

4 SPECIFICATION OF THE R2-R2 TRANSIT CALL-HANDLING PROCESS




12 AWAIT CUT IN BWD MFC
13 RETRANSMIT A3 BWD SIGNAL
14 AWAIT CUT IN FWD SIGNAL
15 AWAIT GROUP II FWD SIGNAL
16 AWAIT GROUP II FWD SIGNAL
17 AWAIT CUT IN A3 BWD SIGNAL
18 GROUP II SIGNAL SENT FWD
ANALYSE BACKWARD SIGNAL b
GROUP B SIGNAL
B1, B6, B7
B2-B5, B8-B15
19 RETRANSMIT SUCCESS SIGNAL
Z = B1, A6/B6 OR B7
NOTE 8: THE $\bar{j}_O$ SIGNAL MUST BE IGNORED WITHIN 75 ms OF ENTERING STATE 19
20 WAIT FOR ANSWER
21 CONVERSATION
22 AWAIT 'B' RE-ANSWER
ICJ
OGJ
MFC (NT)
MFC REC
MFC (A3)
MFC (GPII)
MFC (Z)
CF
U/S
TO
TI
CB
B
p 4


INTERRUPTION OF NORMAL SEQUENCE BY BACKWARD MFC SIGNAL
ANALYSE BACKWARD SIGNAL
GROUP A SIGNAL
ANY OTHER A * SIGNAL
23 START SENDING A3 PULSE
NOTE 9: SEE NOTE 3, p 2
NOTE 10: THE ASTERISKS IMPLY THAT THE STATES OF THE MFC RECEIVERS ARE IGNORED DURING STATES 23, 24 AND 25
100ms
24 MIDDLE OF A3 PULSE
200ms
25 END OF A3 PULSE
100ms
ANALYSE STATES OF MFC RECEIVERS
CLEAR FORWARD SEQUENCE
CALL SPLITS
CLEAR BACK SEQUENCE
CLEAR-DOWN FOLLOWING TIME-OUT IN INCOMING MFC SEQUENCE
26 START A4/A15 PULSE
SEE NOTE 6 p.2
PREPARE TO SEND Z = A4 OR A15
CLEAR-DOWN FOLLOWING TIME-OUT IN OUTGOING MFC SEQUENCE
27 INFORMATION TONE
eg. BUSY TONE OR RECORDED ANNOUCEMENT
INFO TONE
GENERAL "UNSUCCESSFUL" SEQUENCE IN WHICH THE REASON FOR THE LACK OF SUCCESS IS EXPLAINED VIA A BACKWARD MFC SIGNAL, eg A4, A15, B2-B5, B8-B16
28 AWAIT RELEASE-GUARD
NOTE 11: THE LINE CONDITION J_0 OR $\bar{J}_0$ IS IGNORED DURING THIS STATE
250 ms
29 OGJ OUT-OF-SERVICE
FAULT
30 OGJ BLOCKED
31 OGJ IDLE
OGJ SEIZED FOR NEW CALL AS A CONSEQUENCE OF THE PATH SEARCH PROCEDURE FOR THE NEW CALL, SHOWN BETWEEN STATES 4 AND 6 (p.2)
SEIZURE

Discussion

M. WIZGALL, Germany : *You have pointed out in your paper that the SDL is a good tool to describe and document switching systems. Regarding the simulation of a switching system, you mentioned that it would be good to describe both the real systems and the model by means of SDL. This would also make it easily possible to compare reality and model and to enhance the speed to construct and to prove simulation models. As you have written a simulation programme using SDL, could you compare your method with the other ones regarding time to construct and to debug a simulation programme. Which simulation language fits well with SDL.*

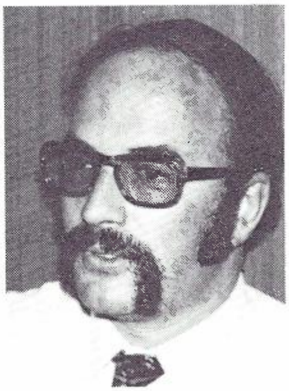
P. GERRAND, Australia : *Thank you for your question. The time taken to construct and debug a simulation programme depends upon the abilities of the designers and programmers, and the quality of the computer simulation language used, as well as the intelligibility and availability of the system description. For this reason, it is possible for an ad hoc simulation to be constructed and debugged as efficiently as a more systematic method, provided that the system description is both intelligible and complete.*

For instance, I have observed that the time required to construct and debug conventional simulation models may range between 4 man-months and 4 man-years, for similar switching systems, depending upon the designers, and programmers' abilities and experience.

The important advantage of using the SDL as the system description for the simulation study, is that this system description can become available to the traffic engineers at a much earlier stage in the system design, and hence the traffic engineers can make a more valuable contribution to the system design, by comparing the effect of different software design choices. My experience in Madrid was that, starting with Level 2 CSTDs, and following discussions with the software designers concerning their intentions, we were able to complete the Level 3 diagrams needed for a simulation model in 2 man-months; and this occurred many months in advance of the coding of the real system.

In answer to your second question, the use of the SDL in a Level 3 system description does not prejudice the choice of simulation language used. My own experience was with SIMSCRIPT, but SIMULA or several other even-by-event simulation languages would have been equally suitable.

BIOGRAPHY



PETER GERRAND is a Senior Engineer in the Network Studies Section, Switching and Signalling Branch, of Telecom Australia's Research Laboratories. He graduated B.Eng. from the University of Melbourne and M.Eng.Sc. from Monash University, supported by an engineering cadetship with the then Postmaster-General's Department. Following two years' military service, he joined the Switching and Signalling Branch as an Engineer in March 1971, and has been working on various aspects of specifying, designing and dimensioning stored-program controlled switching systems ever since. Since 1973 he has been active as the Australian nominated expert on the CCITT Study Group XI's Working Party on Specification and Description Languages for SPC switching systems.

In the two year period from November 1974 to October 1976, Mr. Gerrand was given leave to gain wider industrial experience, working as a consultant to Standard Eléctrica S.A. in Madrid. During this period he was in turn released by Standard Eléctrica for short periods to enable him to represent Telecom Australia at CCITT Study Group XI meetings at Geneva.

Mr. Gerrand's current work activities include the development of a general exchange simulator and a computer graphics aid for the specification of new subscribers' communication facilities.

A Mathematical Model of Telephone Traffic Dispersion in some Australian Metropolitan Networks

A. W. DUNSTAN
Telecom Australia, Brisbane, Australia

ABSTRACT

Analysis of measurements of point to point traffics in telephone networks in some Australian State capitals leads to a form of gravity model in which the explanatory variables are radial distance and telephone service density. The model is extended to estimate "own exchange" traffic and to include an "adjacency factor". A way is suggested of forecasting point to point traffic flows incorporating the influence of exchange parameters as well as measurements of initial point to point traffic flows.

1. INTRODUCTION

Metropolitan telephone networks in Australian State Capitals contain many exchanges linked by a complex junction network in which a considerable number of direct and alternate routes can be provided.

To dimension such a network for economical use of plant, reliable estimates of point to point traffic flows are needed. A programme of traffic dispersion measurements has been under way for some years to provide these estimates. The measurement programme makes the implicit assumption that traffic dispersion is a reasonably stable characteristic of the originating exchange. Forecasts of future point to point traffic are made starting from that assumption.

In the analysis of economic and social interaction between groups of people, mathematical models known as gravity models have often been used to explain the volume of trade or travel etc. The family of models explains the intensity of interaction as being directly proportional to the size of the origin group and to the size of the destination group and inversely proportional to some power of the distance between the two groups. The name gravity model is applied because of the resemblance to the mathematical form of the law of physical gravitation.

A pilot study(1) gave reason to think that a gravity model is appropriate to explain the dispersion of telephone traffic from an origin exchange to various destination exchanges. However a model using only sizes and distance as the explanatory variables may significantly underestimate traffic in some cases and overestimate in others. Demographic factors also appeared to have an important influence on traffic dispersion.

This paper develops a general model for estimating point to point traffics directly from the exchange parameters.

2. NETWORK TRAFFIC FORECASTING

Before attempting to forecast point to point traffic flows in a future network it is necessary to estimate for each exchange:-

The area to be served (and hence the exchange location)

The number of services to be connected

Average calling and terminating traffic rates per service.

From that information the future originating and terminating traffic at each exchange can be estimated.

Kruithof(2) developed his method of double factors which starts with a given set of initial point to point traffics

X_{ij} = Traffic originating at exchange i and terminating at exchange j

and transforms these into a new set of point to point traffics

$$X'_{ij} = p_i\, q_j\, X_{ij}$$

where X'_{ij} = new traffic originating at exchange i and terminating at exchange j

p_i = a constant associated with exchange i as an origin

q_j = a constant associated with exchange j as a destination.

The method consists of finding the values of the factors p_i and q_j which will cause the following constraints to be satisfied:

$$\sum_j X'_{ij} = O_i$$

$$\sum_i X'_{ij} = T_j$$

$$\sum_i O_i = \sum_j T_j = Y$$

where O_i = Total traffic originating at exchange i

T_j = Total traffic terminating at exchange j

Y = Total traffic originating and terminating in the network

The method appears to work very satisfactorily for short run forecasts. However it produces a final solution uniquely determined by the initial pattern of traffic flows and is in no way able to take account of demographic changes. This weakness is most clearly apparent in forecasting traffic flows to and from new exchanges.

Bear and Seymour(3) describe a method where, in the absence of detailed measurements of point to point traffics, these can be estimated from a gravity model. These are used as the starting point for a forecast using the method of double factors.

3. COEFFICIENT OF PREFERENCE

In order to analyse the dispersion of point to point traffic from exchanges of different size, location, traffic rate etc., a dimensionless measure is needed for the tendency of an originating exchange to direct traffic to a particular destination exchange.

For each origin i and destination j, a coefficient of preference can be defined as:

$$K_{ij} = \frac{X_{ij}}{O_i} \cdot \frac{Y}{T_j} \quad \text{..........(1)}$$

where K_{ij} = the coefficient of preference which origin exchange i exhibits for calling exchange j

In some early work it was implicitly assumed that the coefficients of preference were the same for each destination and for all origins, and that they remain constant as the network develops. It is now widely accepted that the first assumption is not generally true and Bear[4] points out that the assumption that the coefficients of preference remain constant as a network of exchanges develops, can lead to unacceptable predictions.

This paper investigates how coefficients of preference vary throughout metropolitan networks and explains this variation in terms of exchange parameters using a mathematical model of the form:

$$K_{ij(E)} = A_0 \, (A_1)^W \left\{\frac{A_2}{D_{ij}}\right\}^{\left\{\frac{1}{A_3 \, P_i}\right\}^{A_4}} \quad \text{..........(2)}$$

where A_0, A_1, A_2, A_3, A_4 are constants

$K_{ij(E)}$ = estimated value of K_{ij}

W = 1 if the destination exchange is adjacent to the originating exchange and is further from the centre of the city and has less than 2 000 services (otherwise W = O)

D_{ij} = distance from exchange i to exchange j in kilometres

p_i = telephone services per hectare in the service area of originating exchange i

The model is not expected to explain every cause of variation in coefficients of preference so at the end of the analysis the original observed coefficients of preference are transformed by the model into residual coefficients of preference defined as:

$$K_{ij(R)} = \frac{K_{ij(O)}}{K_{ij(E)}} \quad \text{..........(3)}$$

where $K_{ij(R)}$ = residual coefficients of preference

$K_{ij(O)}$ = observed coefficient of preference

4. THE SAMPLE ANALYSED

In order to reduce the data handling problem to manageable proportions a sample of the available measurements of traffic dispersion has been selected for analysis. The sample was chosen to give a large coverage of the Brisbane telephone network by selecting recent measurements at twenty six different origin exchanges in a representative range of locations and sizes. Five earlier readings at four of these locations were also included for comparison.

Seventeen Perth network origins were selected in a representative range of sizes and locations. At the time of analysis only five Adelaide network readings were available and all were included.

The Appendix contains data on which this study is based. For each origin, base data is shown in Table I. An illustrative sample for two origin exchanges in the Brisbane network is shown in Table II.

5. DISTANCE EFFECT

In gravity models it is usual to use the travel route distance between the two centres under consideration to account for the fall in preference as distance increases. Of the 945 point to point traffics which were analysed in Brisbane, fifty two were identified where the route distance was appreciably greater than the point to point distance. In most (but not all) of these cases the use of route distance instead of point to point distance in a gravity model would have given a better agreement between the observed coefficient of preference and one estimated from a gravity model.

However it was decided to carry out analysis on the basis of point to point distance, because

(a) in most cases this was little different from the route distance;

(b) data handling is greatly facilitated;

(c) it is possible to include a correction later for the difference in distance (see paragraph 10)

6. PERCEPTION OF DISTANCE

Using a single value of distance exponent in a gravity model to describe telephone users calling patterns implies that callers all through the network share a common scale of perception of the distance of the other telephone services they call and that they exhibit the same "preference slope" favouring nearby telephone services.

The gravity model for that would be

$$K_{ij} = A_0 \left\{\frac{A_2}{D_{ij}}\right\}^n \quad \text{..........(4)}$$

where A_0 and n are constants, which are the same for all originating exchanges and A_2 is a scale factor of distance.

The fifty three sets of data were analysed to estimate a set of coefficients A_i and n_i for each individual originating exchange i. These were far from constant. Using mean values of A_0 and n in a simple gravity model like that used by Bear and Seymour[3] would, in the Australian networks studied, give rise to errors up to 300% in some cases. We must give up the assumption of a common "preference slope" and consider a family of gravity models of the form

$$K_{ij(O)} = A_i \left\{\frac{1}{D_{ij}}\right\}^{n_i} \quad \text{..........(5)}$$

where A_i = a constant for originating exchange i

n_i = a constant for originating exchange i

Comparing equations (4) and (5) shows that testing equation (5) for common scale of perception of distance requires testing the hypothesis that

$$A_i = A_0 \, (A_2)^{n_i} \quad \text{..........(6)}$$

Regression of the fifty three sets of the observed values of A_i on n_i yielded the equation

$$A_i = 0.556(20.79)^{n_i} \quad \text{..........(7)}$$

with a correlation coefficient of 0.99. The coefficients A_0 and A_2 for the model are thus estimated to be:

$$A_0 = 0.556$$

$$A_2 = 20.79$$

The concept of a common scale of perception of distance by all telephone users, is thereby strongly supported. This perception scale uses a unit distance of about twenty one kilometres. The values of A_i and n_i are shown in Table I in the Appendix.

7. PREFERENCE SLOPE

The preference slope (ie the exponent of distance n_i) varied over the range 0.14 to 2.4.

This analysis confirmed the pilot study[1] suggestion that distance exponent is a function of telephone service density. The equation

$$n_i = \left\{\frac{1}{1.113\ p_i}\right\}^{0.399} \qquad \text{..........(8)}$$

where p_i = telephone services per hectare at exchange i,

fitted the data with a correlation coefficient of 0.76. Model coefficients A_3 and A_4 of equation (2) are estimated to be

$$A_3 = 1.113$$

$$A_4 = 0.399$$

Values of p_i are listed in Table I in the Appendix.

SELF DISTANCE

Telephone callers in Australia are not thought to give any weight to whether the number they are calling is connected to their own exchange or to a different one.

However due to the nearness of "own exchange" subscribers the traffic to "own exchange" is usually quite high. Values ranging from 6.5% to 42.5% of originated traffic were found in the data.

Rewriting equation (4) for own exchange traffic allows "self distance" to be estimated as

$$D_{ii} = \left\{\frac{A_i}{K_{ii}}\right\}^{\frac{1}{n_i}} \qquad \text{..........(9)}$$

where D_{ii} = imputed "self distance"

and K_{ii} = coefficient of preference for "own exchange" traffic

The estimated values of D_{ii} are shown in Table I in the Appendix.

As was suggested by Bear[4] these imputed self distances are related to the telephone service density at the originating exchange. The equation

$$D_{ii} = \left(\frac{1}{1.779\ p_i}\right)^{0.6515} \qquad \text{..........(10)}$$

fitted the imputed "self distances" with a correlation coefficient of 0.69.

Note: In the pilot study[1] for nine origins D_{ii} was explained in terms of exchange area but with larger amount of data available this was rejected in favour of area/services i.e. the inverse of telephone service density.

8. ADJACENCY EFFECT

After analysing for the effects of distance and telephone service density the effects of adjacency were examined. Exchanges were classed as adjacent if they have a common boundary but not adjacent if the exchange areas meet only at a point or are separated by a physical barrier such as a river.

Of the 1 568 point to point traffics analysed, 180 involved pairs of exchanges which are adjacent. These readings are indicated by the letter "A" following the observed coefficient of preference $K_{ij(O)}$ in Table II in the Appendix; if the destination is also further from the centre of the city the letters "AF" are printed and if the destination exchange also has less than 2 000 working services, the letters "AFS" are printed. A residual factor was calculated as the ratio of the observed coefficient of preference to the coefficient estimated from the model in equation (2) using coefficients A_0, A_2, A_3, A_4 as estimated above and setting A_1 to unity. There are twenty five observations marked "AFS". For these the geometric mean residual factor is 2.42. For the other 160 cases of adjacency the geometric mean residual factor = 1.17.

This finding estimates the coefficient of equation (2) which accounts for adjacency effect to be:

$$A_1 = (2.42)^w \qquad \text{..........(11)}$$

where A_1 = adjacency factor

w = 1 if AFS occurs,

otherwise $w = 0$

Though the adjacency effect only applies to small parcels of traffic in Australian cities studied, it operates most commonly in rapidly developing areas on the outer fringe.

The most important inference of the adjacency effect is that the strong preference observed for short haul calls to such areas should not be expected to persist when the new area becomes more closely populated.

Two examples of successive measurements involving adjacency effects occur in the data. The residual factors for these cases are shown below.

Mt Gravatt - Eight Mile Plains

Date	Observed Factor	Factor from Equation (11)
1968	2.6	2.42
1974	2.33	1.0

Sunnybank - Eight Mile Plains

Date	Observed Factor	Factor from Equation (11)
1968	2.64	2.42
1970	2.72	2.42
1974	1.79	1.0

Note: Eight Mile Plains has exceeded 2 000 lines since June 1973. A drop in adjacency effect occurred but by late 1974 was not as large as predicted by equation (11).

9. TESTING FOR BIASED ESTIMATION

The functional form of the model in equation (2) is not suitable for multiple regression techniques so it was tested using the individually derived coefficients to calculate residual coefficients of preference and the mean value overall of those coefficients.

A value of 1.108 was obtained for the mean residual coefficient. Coefficient A_0 was adjusted to remove this bias. Because the adjacency factor, A_1, is derived from residual coefficients it does not contain the bias and so it has to be adjusted in the opposite sense when A_0 is adjusted. The final model equation becomes:

$$K_{ij(E)} = 0.616\ (2.18)^{W} \left\{\frac{20.79}{D_{ij}}\right\}^{\left\{\frac{1}{1.113\ p_i}\right\}^{.399}} \qquad ....(12)$$

$$\left\{\text{Note: when } i = j,\ D_{ii} = \left\{\frac{1}{1.779\ p_i}\right\}^{.6515}\right\} \qquad ....(13)$$

10. RESIDUAL COEFFICIENTS

The model in equations (12) and (13) has been used together with equation (3) to calculate for each observation a residual coefficient of preference. The mean value of $K_{ij(R)}$ is unity and its standard deviation is 0.67.

The model equation contains the influence of distance, telephone service density and adjacency. All other influences such as the ratio of travel route distance to point to point distance and also the unreproducible variability of telephone user behaviour, are contained in the residual coefficient of preference.

This separate identification of the demographic variables in the model provides a means of taking account of those influences which is not available by using a simple gravity model or the Kruithof method of double factors.

11. APPLICATION OF THE MODEL TO NETWORK FORECASTING

To make a forecast of point to point traffic flows in a network the procedure is as follows:-

For each exchange which will exist in the future network -

(1) By existing processes, estimate the area it is to serve and its location, the number of services to be connected, the average calling and terminating traffic rates per service (and hence the total originating traffic O_i and terminating traffic T_j).

(2) From the model (equations 12 and 13) calculate $K_{ij(E)}$

(3) From records, read the residual coefficient of preference $K_{ij(R)}$.

(4) Calculate the point to point traffic X_{ij} as

$$X_{ij} = (K_{ij(R)})\ (K_{ij(E)})\ \left(\frac{T_j}{Y}\right)\ O_i \qquad ....(14)$$

Note: If no measurement is available, set

$$K_{ij(R)} = 1$$

(5) Continue for all destinations for all origins

(6) As the traffic matrix thus produced will generally not be balanced, i.e. the sum of the traffic elements will not equal the sum of the originating and terminating traffics, use the Kruithof method of double factors to satisfy those constraints.

12. CONCLUSION

The larger scale analysis reported here generally confirms the form of the model indicated in the earlier pilot study[1]. The following table compares the earlier model equation obtained from nine sets of observations in one city with that now derived from fifty three sets in three cities.

Coefficient	Pilot Study	This Study
Constant	0.68	0.616
Unit Distance	13.3 km	20.79 km
Exponent of Distance	$\left(\frac{1}{1.06\ P_i}\right)^{.52}$	$\left(\frac{1}{1.113\ P_i}\right)^{.399}$
Adjacency Factor	$\left(\frac{7132}{L_{ij}}\right)^{.44}$ (= 2.37 for 1 000 lines)	2.18 for a destination exchange smaller than 2 000 lines which is further from the city centre than the originating exchange. The factor is unity for other cases.
Self Distance	$\left(\frac{\text{Exch Area}}{2850}\right)^{2}$	$\left(\frac{\text{Exch Area}}{1.779\ (\text{Lines})}\right)^{.6515}$

The differences are attributed to the greater information available from the large sample than from the small one.

Samples from the three cities vary from the overall somewhat but there is not sufficient difference to attribute this to anything but the reduced size of the sample.

The method of forecasting described has the advantage that a logical system exists for introducing demographic factors and new exchanges are easily added. The mathematical model uses explicit algorithms based on stated parameters, to estimate future traffic. The Kruithof method of matrix balancing is used only to make small adjustments to the estimates.

The relationships between telephone service density and both the exponent of distance and the "self distance" and the relationship between destination size and adjacency effect observed in this study, were derived from a fairly large group of observations made over a limited span of time. Whether individual exchanges, as they grow in the future, will take on in turn the different characteristics observed for larger and larger exchanges, will require observations over an extended period of time. In the meantime it seems prudent to be prepared for some substantial changes.

The model structure is expected to have general application to telephone networks. However the scale factor of perception of distance is likely to differ where the unit fee call area has a different size and where population is spread more uniformly.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of various colleagues who assisted with data gathering analysis and criticism. He also wishes to acknowledge the permission granted by the Australian Telecommunications Commission to publish this paper.

BIBLIOGRAPHY

(1) A. Dunstan 'A Mathematical Model of Telephone Traffic Flow in the Brisbane Area'. Technical Paper of Queensland Division of Institution of Engineers Australia March 1973 Vol 14 No.2.

(2) J. Kruithof 'Telefoonverkeerscrekening'. De Ingenieur (E. Electrotechmek 3) Vol. 52 No.8, 19 February 1937.

(3) D. Bear and C. A. Seymour 'A Traffic Prediction Model for a Telephone Exchange Network'. Seventh International Teletraffic Congress Stockholm 1973.

(4) D. Bear 'Some Theories of Telephone Traffic Distribution : A Critical Survey'. Seventh International Teletraffic Congress Stockholm 1973.

APPENDIX

BRISBANE ORIGINS	DATE	A_i	n_i	r_i	p_i	D_{ii}
ASHGROVE	July 1969	4.23	0.72	0.7	3.93	0.1
SUNNYBANK	July 1964	29.9	1.51	0.84	0.57	0.98
SUNNYBANK	July 1970	10.2	0.98	0.77	1.55	0.42
ASPLEY	May 1969	8.34	0.89	0.83	0.82	1.23
TINGALPA	May 1970	23.3	1.4	0.69	0.35	2.15
ASCOT	April 1970	2.08	0.47	0.49	2.21	1.34
CHAPEL HILL	July 1967	6.48	0.87	0.69	1.45	0.45
VALLEY	Sept 1970	1.27	0.26	0.42	9.18	0.88
W'GABBA	April 1971	1.5	0.32	0.56	6.51	0.42
ALBION	Jan 1973	1.63	0.3	0.5	9.97	0.05
BROOKFIELD	Dec 1973	13.1	1.06	0.5	0.08	3.74
BULIMBA	Dec 1974	1.9	0.42	0.49	3.85	0.28
CHAPEL HILL	May 1972	8.61	1	0.83	2.22	0.36
CHERMSIDE	Dec 1974	6.22	0.86	0.75	6.46	0.25
CLEVELAND	May 1971	1600	2.42	0.75	0.32	0.73
EDISON	June 1975	1.07	0.2	0.68	52.9	0.19
INALA	Sept 1969	42.2	1.43	0.77	0.19	3.39
INALA	Nov 1974	10.7	0.94	0.76	1.03	0.74
IPSWICH	Feb 1975	163	1.72	0.93	1.19	0.45
MT GRAVATT	May 1968	9.51	1.01	0.82	1.27	0.96
MT GRAVATT	Nov 1974	8.77	1.02	0.89	4.07	0.37
NEW FARM	Oct 1975	1.8	0.44	0.81	9.55	0.02
NEWMARKET	Nov 1972	2.18	0.41	0.5	5.83	0.23
NUNDAH	Dec 1971	4.09	0.68	0.62	4.08	0.27
SANDGATE	Feb 1973	25.1	1.28	0.89	1.75	0.48
SLACKS CK	July 1972	149	1.73	0.77	0.18	2.46
SOUTH	Nov 1973	1.24	0.25	0.51	7.25	0.82
STRATHPINE	Sept 1974	38	1.35	0.92	0.33	1.95
SUNNYBANK	Dec 1974	15.6	1.19	0.91	2.16	0.35
WYNNUM	April 1975	48	1.57	0.69	2.43	0.19
FERNY HILLS	Sept 1974	35.9	1.45	0.78	0.27	1.51
PERTH ORIGINS						
APPLECROSS	Dec 1975	4.71	0.74	0.68	5.46	0.03
BASSENDEAN	Dec 1975	10.5	1.01	0.7	1.37	0.81
BATEMAN	Dec 1975	11.5	1.01	0.79	0.28	1.96
BULWER	Dec 1975	2.35	0.25	0.48	9.04	0.02
CANNINGTON	Dec 1975	6.02	0.75	0.42	2.63	0.99
CENTRAL	Dec 1975	1.05	0.18	0.52	13.1	0.28
CITY BEACH	Dec 1975	10.8	1.09	0.65	1.49	0.18
MAYLANDS	Dec 1975	1.28	0.14	0.2	4.28	0.19
MIDLAND	Dec 1975	34	1.34	0.85	0.42	2.35
VICTORIA PK	Dec 1975	1.23	0.15	0.19	4.98	0.26
KALAMUNDA	Dec 1975	183	1.83	0.82	0.46	0.86
KELMSCOTT	Dec 1975	65.1	1.44	0.81	0.55	0.88
GOSNELLS	Dec 1975	23.2	1.1	0.78	0.26	2.90
GREENMOUNT	Dec 1975	38	1.33	0.74	0.54	0.61
TUART HILL	Dec 1975	5.71	0.76	0.58	3.46	0.19
WEMBLEY	Dec 1975	3.88	0.71	0.6	3.85	0.06
SPEARWOOD	Dec 1975	57.8	1.5	0.76	0.56	1.15
ADELAIDE ORIGINS						
NORWOOD	June 1975	3.94	0.76	0.84	6.03	0.34
SEMAPHORE	June 1975	8.66	0.93	0.73	8.01	0.05
ST PETERS	June 1975	1.81	0.41	0.53	10.2	0.28
BLACKWOOD	June 1975	4.93	0.72	0.6	2.07	0.12
W ADELAIDE	June 1975	1.24	0.21	0.34	6.19	0.71

LEGEND:

Deriving equation (5) $K_{ij(0)} = A_i \left(\frac{1}{D_{ij}}\right)^{n_i}$ by regression

A_i = constant for origin exchange i
n_i = (neg) exponent of distance for origin exchange i
r_i = correlation coefficient for origin exchange i
p_i = telephone services per hectare for origin exchange i
D_{ii} = imputed self distance for exchange i

TABLE 1

IDENTITY AND PARAMETERS OF ORIGIN EXCHANGES

DESTINATION	MT GRAVATT 1968 $K_{ij(O)}$	MT GRAVATT 1968 $K_{ij(E)}$	MT GRAVATT 1968 $K_{ij(R)}$	MT GRAVATT 1974 $K_{ij(O)}$	MT GRAVATT 1974 $K_{ij(E)}$	MT GRAVATT 1974 $K_{ij(R)}$	SUNNYBANK 1964 $K_{ij(O)}$	SUNNYBANK 1964 $K_{ij(E)}$	SUNNYBANK 1964 $K_{ij(R)}$	SUNNYBANK 1970 $K_{ij(O)}$	SUNNYBANK 1970 $K_{ij(E)}$	SUNNYBANK 1970 $K_{ij(R)}$	SUNNYBANK 1974 $K_{ij(O)}$	SUNNYBANK 1974 $K_{ij(E)}$	SUNNYBANK 1974 $K_{ij(R)}$
EDISON	0.86	1.14	0.75	0.81	0.91	0.89	0.53	1.14	0.47	0.74	0.93	0.8	0.71	0.88	0.81
CENTRAL	0.74	1.14	0.65	0.87	0.91	0.96	0.7	1.14	0.61	0.85	0.93	0.91	0.91	0.88	1.03
PADDINGTON	1.58	0.98	1.61	0.7	0.83	0.85	0.38	1	0.38	0.54	0.85	0.63	0.58	0.82	0.71
ASHGROVE	0.42	0.82	0.51	0.46	0.74	0.62	0.52	0.83	0.62	0.54	0.75	0.72	0.43	0.73	0.58
THE GAP	< MIN	0.72	0.04	0.5	0.68	0.73	0	0	0	0.56	0.69	0.81	0.22	0.68	0.33
SOUTH	1.05	1.29	0.82	0.99	0.98	1.02	1.13	1.35	0.84	1.03	1.04	0.99	1.02	0.98	1.04
SALISBURY	1.08 A	1.96	0.55	1.12 A	1.27	0.88	3.33 A	4.59	0.72	3.88 A	2.37	1.63	2.78 A	2.01	1.38
YERONGA	1.62	1.89	0.86	1.15	1.25	0.92	1.56	2.39	0.65	1.52	1.53	1	1.47	1.37	1.08
MT GRAVATT	8.64 *	13.7	0.63	5.72 *	6.57	0.87	2.42 A	4.12	0.59	2.09 A	2.21	0.95	2.65 A	1.88	1.41
E.M. PLAINS	4.61 AFS	4.21	1.09	2.68 AF	1.28	2.1	7.32 AFS	6.59	1.11	4.44 AFS	3.89	1.14	2.56 AF	1.58	1.62
ACACIA RIDGE	0	0	0	1	1.07	0.94	0	0	0	3.84 AFS	5.28	0.73	3.74 AFS	4.45	0.84
VALLEY	0.73	1.05	0.69	0.71	0.86	0.82	0.68	1.01	0.67	0.74	0.86	0.86	0.62	0.82	0.75
MITCHELTON	0.63	0.7	0.9	0.44	0.67	0.65	0.65	0.66	0.98	0.63	0.65	0.98	0.52	0.64	0.82
NEWMARKET	0.53	0.84	0.62	0.66	0.75	0.88	0.34	0.8	0.42	0.65	0.74	0.88	0.41	0.72	0.56
LUTWYCHE	0.5	0.83	0.6	0.46	0.74	0.63	0.24	0.75	0.32	0.6	0.7	0.86	0.39	0.69	0.57
NEW FARM	0.89	1.19	0.75	0.81	0.93	0.87	0.51	1.13	0.45	0.76	0.93	0.82	0.92	0.88	1.04
CHERMSIDE	0.55	0.7	0.79	0.34	0.67	0.5	0.36	0.62	0.58	0.41	0.62	0.67	0.35	0.62	0.57
FERNY HILLS	0	0	0	0.49	0.6	0.82	0	0	0	0	0	0	0.69	0.58	1.19
ALBION	0.44	0.88	0.5	0.34	0.77	0.44	0.47	0.8	0.59	0.44	0.73	0.6	0.4	0.72	0.55
ALPLEY	0.46	0.59	0.77	0.35	0.6	0.58	0.53	0.51	1.04	0.18	0.54	0.34	0.29	0.55	0.53
NUNDAH	0.71	0.74	0.96	0.43	0.69	0.62	0.60	0.64	0.93	0.65	0.63	1.02	0.57	0.63	0.89
ASCOT	0.34	0.92	0.37	0.5	0.79	0.63	0.51	0.81	0.63	0.51	0.74	0.69	0.49	0.72	0.68
SANDGATE	0.44	0.52	0.85	0.43	0.55	0.78	< MIN	0.42	0.55	0.48	0.48	1	0.29	0.49	0.58
NUDGEE	0.42	0.65	0.65	0.42	0.64	0.65	< MIN	0.55	0.28	0.37	0.57	0.65	0.45	0.57	0.79
BALD HILLS	< MIN	0.52	0.49	< MIN	0.55	0.19	< MIN	0.43	0.12	< MIN	0.48	0	0.63	0.5	1.26
PINKENBA	0.43	0.86	0.5	0.32	0.76	0.42	< MIN	0.72	0.04	0.45	0.69	0.65	0.74	0.68	1.09
ALBANY CREEK	0	0	0	0.64	0.56	1.16	0	0	0	0	0	0	0.69	0.51	1.36
ZILMERE	0	0	0	0.55	0.61	0.91	0	0	0	0	0	0	< MIN	0.55	1.01
TOOWONG	0.51	1.05	0.49	0.59	0.86	0.69	0.26	1.19	0.21	0.27	0.96	0.29	0.4	0.91	0.44
INALA	0.39	1	0.39	0.63	0.84	0.75	0	0	0	0.93	1.32	0.71	1.09	1.2	0.91
JINDALEE	< MIN	0.89	0	0.79	0.76	1.03	0	0	0	< MIN	0.98	0.41	1.18	0.94	1.25
CHAPEL HILL	0.42	0.92	0.46	0.54	0.79	0.68	0.53	1.13	0.47	0.68	0.92	0.73	0.66	0.88	0.75
SHERWOOD	0.62	1.18	0.52	0.83	0.93	0.9	1.12	1.75	0.64	0.83	1.24	0.66	0.92	1.14	0.8
DARRA	0.79	0.94	0.84	0.9	0.81	1.12	0.99	1.5	0.66	1.23	1.12	1.1	0.44	1.04	0.43
BROOKFIELD	0.91	0.69	1.32	0.19	0.66	0.29	2.35	0.76	3.07	< MIN	0.71	0.91	0.76	0.7	1.09
IPSWICH	0.4	0.42	0.96	0.26	0.48	0.54	0	0	0	0.38	0.47	0.81	0.28	0.49	0.57
BUNDAMBA	0.39	0.48	0.8	0.28	0.53	0.53	0	0	0	< MIN	0.54	0.69	0.54	0.55	0.98
REDCLIFFE	0.35	0.39	0.88	0.3	0.47	0.64	< MIN	0.3	0.74	0.49	0.38	1.29	0.28	0.4	0.7
PETRIE	0.51	0.42	1.23	0.21	0.48	0.44	< MIN	0.33	0.09	< MIN	0.41	1	0.42	0.43	0.99
STRATHPINE	0.62	0.47	1.32	0.17	0.52	0.33	< MIN	0.38	0.35	< MIN	0.45	0.99	0.35	0.47	0.74
CLEVELAND	1.18	0.67	1.76	0.94	0.65	1.45	< MIN	0.57	0.08	< MIN	0.59	0.32	0.53	0.59	0.9
BEENLEIGH	< MIN	0.58	0	1.57	0.59	2.65	3.14	0.6	5.21	< MIN	0.61	1.27	1.25	0.61	2.06
SLACKS CREEK	0.7	1.09	0.64	1.6	0.88	1.81	1.52 AFS	3.45	1.31	5.04	1.17	4.3	2.49	1.08	2.3
W'GABBA	1.53	1.36	1.13	1.42	1.02	1.4	1.18	1.36	0.87	1.21	1.05	1.15	1.42	0.98	1.45
SUNNYBANK	2.19	2.45	0.9	1.92	1.47	1.31	22.1 *	23.6	0.94	14.2 *	12	1.18	11.2 *	9.69	1.16
BULIMBA	1.13	1.34	0.84	1.07	1	1.07	0.76	1.22	0.62	0.94	0.98	0.96	0.96	0.92	1.04
WYNNUM	0.65	0.84	0.78	0.82	0.75	1.1	0.59	0.69	0.86	0.73	0.67	1.1	0.55	0.66	0.83
COORPAROO	2.73 A	2.04	1.34	1.92 A	1.31	1.47	0.74	1.96	0.38	1.16	1.34	0.86	1.04	1.22	0.86
CAMP HILL	1.6 A	1.88	0.85	1.17 A	1.24	0.94	1.37	1.61	0.85	0.89	1.17	0.75	0.97	1.08	0.9
CAPALABA	1.12 AFS	2.34	0.48	1.21 AFS	1.9	0.64	< MIN	0.94	0.68	< MIN	0.82	1.2	1.08	0.79	1.37
TINGALPA	0.85	1.23	0.69	1.1	0.95	1.16	0.79	1.02	0.77	0.64	0.87	0.74	0.97	0.83	1.17

LEGEND:

$K_{ij(O)}$ = Observed Coefficient of preference which the named origin exchange shows for destination exchange j

$K_{ij(E)}$ = Coefficient estimated from the model, which the named origin would show for destination exchange j

$K_{ij(R)} = \frac{K_{ij(O)}}{K_{ij(E)}}$ = residual coefficient of preference

* is printed when i = j and indicates 'own exchange preference'

< MIN indicates observed traffic from origin i to destination exchange j is less than 0.1E or is less than ½% of traffic originated at i

A is printed if destination exchange is adjacent to origin exchange

AF is printed if destination exchange is also further from the centre of the city than the origin exchange

AFS is printed if the destination exchange also has less than 2 000 services

TABLE II. SAMPLE OF COEFFICIENTS OF PREFERENCE (OBSERVED, ESTIMATED & RESIDUAL) IN TWO BRISBANE NETWORK EXCHANGES

APPENDIX (CONT'D)

Discussion

R. KHADEM, Canada : Your gravity model understates the impact of the "community of interest" in determination of traffic. My question relates to the degree of confidence one can have on a longer term network forecast, especially in circumstances where the "community of interest" itself may be steadily evolving or changing in structure, (or may be the case in Australia). You have hinted at this very point towards the end of your paper. Could you further elaborate, and state over what period you estimate your results are useful.

A.W. DUNSTAN, Australia : I believe that the "coefficient of preference" defined by a number of writers including Daisenberger (paper 341) and myself, is the best measure of the general term "community of interest". I agree that "coefficients of preference" are steadily changing and this is an essential part of my model. This model should be useful over periods during which the values of the coefficients A_0 - A_4 of the model equation do not change too greatly. These "constants" will change but at a much slower rate than the coefficients of preference. Over a period greater than 10 years for example I would expect the "scale of distance" coefficient A_2 to increase somewhat due to increasing mobility of people.

P.A. CABALLERO, Spain : In your model of telephone traffic dispersion radial distance, telephone density and adjacency are the factors influencing the point to point traffic. You also introduce a residual coefficient to take into account other influences, in particular, the ratio of travel route distance to point to point distance. Is this a significant factor? If it is, do you have some data quantifying influence.

A.W. DUNSTAN, Australia : Only a small fraction of cases (15%) involved big difference in route distance versus radial distance. The difference in coefficient of preference is noticeable but is subsumed into residual coefficients of preference.

E. WALDRON, Australia : Your model does not appear to allow for the effect of natural barriers;
e.g. Rivers, Freeways, Lakes, Railway Lines, etc.
I feel sure that these would effect the "perceived distance" and hence suggest the inclusion of these in your model may greatly increase its accuracy.
Please comment on this aspect of natural barriers.

A.W. DUNSTAN, Australia : Metropolitan telephone networks do not appear to contain many absolute natural barriers. The effect of rivers and others which you list appears to be relatively small and adequately explained in terms of a greater route distance caused by the barrier.

As discussed in the answer to Mr. Caballero's question, correction for this effect is included in the residual coefficients.

E. WALDRON, Australia : Since the system you propose implies establishing and maintaining a special set of parameters it appears to me an expensive system compared to an ad hoc guess of dispersion. Such an ad hoc estimate is necessary for a new origin and "nominal" errors can generally be accommodated in practice by the alternate-routed system.
Please comment.

A.W. DUNSTAN, Australia : All of the parameters needed by my model are already recorded in the existing planning processes in my office (see para.11(1) and I would expect these to be available in most administrations. The coefficients of preference are not separate from the base traffic matrix which is usually available. Daisenberger in para 5(6) of Paper 341 shows how to transform one into the other. By using residual coefficients, as I suggest, the only additional parameters to be stored are the five coefficients of the model equation. I do see a virtual necessity in any case to retain records of the network history in the form of matrices of traffic and the corresponding exchange parameters. An informal estimate could certainly be made of the dispersion for a new region. That estimate would take into account, informally, the important factors considered likely to affect the dispersion. My model formalises the process of making the estimate. I take your final remark to be related to the inherently large variability of traffic. I agree this is large and note the results reported in Paper 262 Fig. 4 and Para. 23 for the U.K. This could make "Extreme Value Engineering" suggested by Barnes in Paper 242 of wider application.

R.B. POTTS, Australia : The model expressed by equations (12), (13) involves various parameters calculated from actual traffic.

(i) How is the accuracy to which these parameters are given (e.g. 1.113, 20.79) justified.

(ii) What is the basis for the assumption that these same values can be used for forecasting future traffic.

A.W. DUNSTAN, Australia :
(i) The implied accuracy is not real at best 2 figures are significant.

(ii) 2nd last para. of conclusion deals with this. There is no basis yet but study of secular trend will answer this. Only a short history is available and trends are persuasive rather than definitive.

D. BEAR, U.K. : Have you considered a law of the form

$$K_{ij} = A_0 e^{-nDij}$$

(or similar) as an alternative to equation 4? One form of the gravity model, derived from entropy considerations, suggests a negative exponential rather than an inverse power law of distance. It is interesting that certain Australian road traffic data have been adduced in support of this model.
J.A. and S.G. Tomlin "Traffic Distribution and Entropy" Nature, 220, 974-976, 7th December 1968.

A.W. DUNSTAN, Australia : No. Initially scatter diagrams were produced for many origins on log-log graph paper. The patterns appeared consistent with the ordinary form of gravity model. There is a group of observations of abnormally low coefficients of preference to very distant destinations in my data for which the negative exponential form of model may give an improved explanation. However there are more of the very distant coefficients which appear to be explained by the ordinary gravity model. The negative exponential form would worsen the fit at short distances where traffic flow is high. It is proposed to do some more analysis to check if the C.B.D. may be exerting a "Shielding" effect when it lies between origin and destination (as it often does on very long distance calls).

L. LEE, Canada : Can you give us some indications whether other mathematical models of telephone traffic dispersion have been considered. If not, why not.

A.W. DUNSTAN, Australia : As indicated in the answer to Mr. Bear's question, the data appear to be more consistent with a distance effect function which uses a negative power of distance.

The "preference slope" is a subscribers behaviour characteristic. Variation in this characteristic was observed to be related to several characteristics of the subscribers such as "(i) distance from central business district (ii) subscribers BHCR (iii) percentage of business services (v) telephone density per hectare" which are themselves strongly correlated. Telephone density was chosen for an explanatory variable in the model as it was a demographic variable which was always available and involved no definition problems. In a preliminary study it appeared to be more powerful than BHCR. Further checking will be carried out.

BIOGRAPHY

A.W. DUNSTAN joined the Postmaster-General's Department as an Engineer in January 1950 after graduating from the University of Queensland with the degree of Bachelor of Engineering (Electrical Communications). He subsequently was awarded the degree of Bachelor of Economics from the same University in 1972.

His time as an Engineer was mainly in Exchange Installation but with 10 years in Planning. Currently he is Supervising Engineer, Fundamental Planning in Brisbane.

Some Formulae Old and New for Overflow Traffic in Telephony

C. PEARCE AND R. POTTER

University of Adelaide, Adelaide, Australia

ABSTRACT

The description of telephone traffic by channel occupancies leads, in a situation of repeated overflows, to complicated equations for the joint distribution of the numbers of occupied channels at the different stages involved. We employ a prescription through the times separating consecutive calls in the traffic stream, which enables compound systems to be considered piecemeal. Some specific formulae are derived generalising known formulae and some general questions considered through this approach.

1. INTRODUCTION

The second and subsequent overflow traffics resulting from a first-routed Poisson stream of calls has a non-random character, which produces difficulties in the analysis of stochastic teletraffic systems involving sequential overflows. This problem is compounded by a frequent need to treat not merely simple overflows but pooled overflows resulting from more than one group of channels. Such difficulties are set aside in practice by Wilkinson's "equivalent random" method in which the traffic under consideration at any stage is replaced by a traffic with the same mean and variance which is the first overflow resulting from the input of a Poisson stream to some group of trunks (Ref. 1). However no complete theoretical validation of the robustness of this technique for compound systems appears to exist, although studies of some simple cases have been done by Neovius (Ref. 2), Le Gall (Ref. 3) and others. An exact analysis of the first overflow traffic arising from a Poisson stream is given by Riordan's appendix (Ref. 1), and is used as the basis of the equivalent random method, but no general extension to subsequent overflows is known to the authors. This and more complicated questions such as the relative overflows resulting from pooled non-random traffic appear with a few exceptions to be completely unresolved.

As overflow traffic is non-random, the evolution with timme of the distribution of the number of busy channels at some point of a system from an arbitrary instant depends not only on the number of calls present but also on the time since the advant of the last call; the system has memory. The lack of memory property can be recovered for an arbitrary instant only at the cost of incorporating further information via the joint distribution of the number of busy channels at all previous stages, leading to very complicated equations. Our point of departure lies in specifying the stream not by channel occupancy but through the times separating the advents of successive calls in the stream. The times separating successive calls in an overflow are independently and identically distributed if the input has this property, as was proved by Palm (Ref. 4) and is clear from our subsequent analysis. The description of inter-event times for such a renewal stream is particularly simple, requiring but a single distribution function. The specification is more complicated if at some stage non-random renewal streams are pooled, and for the present we shall not consider that possibility.

Our present object is to provide a theoretical tool useful in the exact analysis of tandem systems with repeated overflows. Our approach is based on the determination of the overflow inter-event time distribution G(t) corresponding to an input distribution function F(t) (and some given number of channels). This enables tandem systems with iterated overflows to be tackled piecemeal, relating each stage only to the immediately preceding one, once we have found expressions for the traffic induced by a given G in a set of N channels. In the following section we determine G in terms of F. The conventional channel occupancy measure of traffic is then related to our G, as is the actual traffic carried by a finite set of channels. With this basic machinery in hand we finally address ourselves briefly to some general questions involving overflow traffic.

The description of a stream of calls by an inter-event time distribution function is standard in queueing theory and some of our basic occupancy formulae (for one and two stages) and formulae connecting input and output streams occur in diverse places in the literature, and are thus certainly not new. However, since we are aiming at a unified methodology and have, we believe, a very simple treatment and formulation, we have seen fit to prove all our results *ab initio*. We wish to demonstrate how the repeated use of a single technique suffices for the derivation of all our basic machinery. Our self-contained treatment is also convenient for the use of certain subsidiary results.

Section 2 derives expressions giving the overflow stream in terms of the offered traffic. This problem was first treated by Palm (Ref. 4) and later by Takács (Ref. 5). In Section 3 the inter-event description of traffic is related to the usual description in terms of the distribution of the number of channels in an infinite set which would be occupied if that traffic were presented to them. This amounts to a study of the G/M/∞ queue. In the case of Poisson offered traffic this problem has been examined by Kosten (Refs. 6,7) and subsequently by Riordan (Ref.1). Our more general case is considered by Takács (see Ref.8) and by Cohen (Ref. 9). The associated problem of the overflow traffic carried by a finite set of trunks is analysed in Section 5. This was studied in the case of original Poisson offered traffic by Brockmeyer (Ref. 10) and in the general case again by Takács and Cohen.

Le Gall (Ref. 3) has considered a renewal stream overflowing from a finite set of trunks and being presented to a second finite set. This corresponds to our work of Section 6 restricted to a first overflow. His distribution of the number of occupied channels at the second stage was found, as with simpler cases treated in the literature, via the joint distribution of the numbers of occupied channels at the two stages.

We adopt without further comment the standard assumption of negative exponential channel holding times and denote the mean holding time by μ^{-1}.

2. RELATION BETWEEN F AND G

Consider an input stream with inter-event time distribution function F(t) offered to a group of N channels. For $0 \leq n < N$ represent by $f_n(t)$ the distribution function for the time separating an epoch when a call joins the group of N channels to find n trunks already occupied and the instant of the first subsequent overflow. Then

$$f_n(t)= \sum_{j=o}^{n+1} \int_o^t \binom{n+1}{j}(1-e^{-\mu y})^{n+1-j} e^{-\mu j y} f_j(t-y)\,dF(y), \quad o \leq n<N,$$

where we interpret $f_N(t)$ as $\delta(t-0)$. If we write $f_n^*(s)$ for the Laplace-Stieltjes transform of $f_n(t)$, that is

$$f_n^*(s) = \int_o^\infty e^{-st}\, df_n(t), \quad \mathrm{Re}\ s \geq 0,$$

then we can express these equations for $n = 0,1,..,N-1$ as

$$f_n^*(s)= \sum_{j=0}^{n+1} \int_0^\infty e^{-st}\binom{n+1}{j}(1-e^{-\mu t})^{n+1-j}e^{-j\mu t}dF(t)f_j^*(s), \quad (1)$$

for $0 \leq n < N$ and $\mathrm{Re}\ s \geq 0$. The solution to these equations is the same as that of the corresponding unrestricted set (2) for which $0 \leq n < \infty$ and there is an imposed supplementary condition $f_N^*(s) = 1$. To solve these equations, define the formal generating function

$$f^*(z) = \sum_{n=0}^{\infty} f_n^*(s)z^n/n!$$

Then on taking generating functions in (2) we have formally, after an interchange of orders of summation

$$f^*(z)=\sum_{j=1}^{\infty} f_j^*(s)z^j \sum_{r=0}^{\infty}\int_0^\infty e^{-st}z^{r-1}\binom{r+j}{r}(1-e^{-\mu t})^r e^{-\mu jt}dF(t)$$
$$+ f_0^*(s)\int_0^\infty e^{-st}(1-e^{-\mu t})\exp(z(1-e^{-\mu t}))dF(t)$$
$$= \int_0^\infty e^{-st}\ d/dz[f^*(ze^{-\mu t})\exp(z(1-e^{-\mu t}))]dF(t).$$

If we set

$$k(z) = f^*(z)e^{-z}$$

with a formal series expansion $k(z)=\sum_{n=0}^{\infty} k_n(s)z^n/n!$, then

$$k(z) = \int_0^\infty e^{-st}(1+d/dz)k(ze^{-\mu t})dF(t),$$

so that on expanding and equating like powers of z we have

$$k_{n+1} = k_n(s)\ [1-\phi(s+n\mu)]/\phi(s+(n+1)\mu).$$

Thus

$$k_n(s) = k_o(s) \prod_{j=1}^{n}[1-\phi(s+(j-1)\mu]/\phi(s+j\mu), \quad (4)$$

and finally the f_n^* may be recovered by inverting (3) as

$$f_n^*(s) = \sum_{j=0}^{n} \binom{n}{j}k_j(s).$$

The undetermined multiplier $k_o(s)$ is fixed by the supplementary condition. In particular, we derive

$$f_{N-1}^*(s) = [\sum_{r=0}^{N-1} \binom{N-1}{r}k_r(s)]/[\sum_{r=0}^{N} \binom{N}{r}k_r(s)],\ \mathrm{Re}\ s\geq 0,$$

and without loss of generality we may take $k_o(s) = 1$ in (4) for the purpose of evaluating this expression. We observe that for $s = 0$, $f^*(z)$ becomes e^z and $k(z)=1$, so that $k_n(0) = \delta_{no}$. The above solution can be verified *post hoc* by substitution in the original equations (1).

The point of the above gymnastics is as follows: the separation of consecutive overflows from the group of channels is the time between two consecutive epochs at which an arriving call finds all N channels occupied, and so has distribution function G(t) given by

$$G(t) = \sum_{j=0}^{N} \binom{N}{j}\int_0^t(1-e^{-\mu y})^{N-j}e^{-\mu jy}f_j(t-y)dF(y).$$

But this is precisely the expression we have given for $f_{N-1}(t)$, as is reasonable physically, since after the arrival of one more call the full set of channels and that with N-1 calls become indistinguishable (except for the overflowing call). Thus the overflow distribution function G(t) has Laplace-Stieltjes transform

$$\psi(s) = [\sum_{r=0}^{N-1} \binom{N-1}{r}k_r(s)]/[\sum_{r=0}^{N} \binom{N}{r}k_r(s)],\quad \mathrm{Re}\ s\geq 0, \quad (5)$$

where the k_r can be evaluated by (4) with $k_o(s)$ taken as unity. This gives $\psi(s)$ in terms of $\phi(s)$.

The above result also gives that the distribution of the non-busy period of a group of N trunks is identical to the inter-overflow distribution of a group of N-1 trunks, a result found by Descloux (Ref. 11) for the special case of Poisson offered traffic.

3. CHANNEL OCCUPANCY DESCRIPTION OF G

Suppose that a renewal stream of calls with inter-event time distribution function G(t) is presented to an infinite number of channels in parallel. The usual measure of the traffic carried by the stream is then the steady-state distribution $\{q_j;\ j\geq 0\}$ in continuous time of the number of busy channels. The analysis is, however, most easily effected through the "imbedded chain" technique of queueing theory, whereby we find the steady-state distribution $\{\pi_j;\ j\geq 0\}$ of the number of busy channels as found by an arriving call. The q-distribution is then found from the π-distribution. Unless G is negative exponential, the distributions π- and q- are in general different. In a practical context in which traffic is presented to a finite number N of channels, the overflow probability is clearly π_N rather than q_N so that the π-distribution is of some interest in its own right. In the common case of random offered traffic the result $\pi_N = q_N$ holds, which has tended to obscure the distinction between the two distributions.

By considering the possible changes for the channels between two consecutive arrival instants, or directly by an appeal to the fundamental ergodic theorem for Markov chains, we have for $j \geq 1$

$$\pi_j = \sum_{m=j-1}^{\infty} \pi_m\int_0^\infty\binom{m+1}{j}e^{-j\mu x}(1-e^{-\mu x})^{m+1-j}dG(x), \quad (6)$$

or, in terms of the probability generating function

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j, \quad |z|\leq 1,$$

$$\pi(z) = \pi_o + \int_0^\infty w[1+(z-1)e^{-\mu x}]dG(x)-\int_0^\infty w(1-e^{-\mu x})dG(x),$$

where $w(z) = z\pi(z)$. Use of the normalising condition $w(1) = 1$ allows this expression to be simplified to

$$\pi(z) = \int_0^\infty w[1+(z-1)e^{-\mu x}]dG(x).$$

Taking n-th derivatives at $z = 1$ then gives

$$\pi^{(n)}(1) = n\pi^{(n-1)}(1)\psi(n\mu)/[1-\psi(n\mu), \quad (7)$$

so that as $\pi^{(0)}(1) = 1$, we have

$$\pi^{(n)}(1)/n! = h_n, \quad n \geq 0, \quad (8)$$

where $h_n = \prod_{m=1}^{n} \psi(m\mu)/[1-\psi(m\mu)]$, $n \geq 0$, and we interpret the empty product h_o as unity. In the simple case where the stream is Poisson with parameter λ, we have $h_n=A^n/n!$, where $A = \lambda/\mu$. $\pi(z)$ may be recovered explicitly in terms of the binomial moments $\pi^{(n)}(1)/n!$ through the Taylor expansion

$$\pi(z) = \sum_{n=0}^{\infty} (z-1)^n\ \pi^{(n)}(1)/n!$$

about centre $z=1$ as $\pi(z) = \sum_{n=0}^{\infty}(z-1)^n h_n$, whence

$$\pi_j = \sum_{n=j}^{\infty} (-1)^{n-j}\binom{n}{j}h_n, \quad j\geq 0.$$

The continuous time q-distribution is easily accessible through the π-distribution. Analogously to (6) we have

$$q_j=\sum_{m=0}^{\infty}\pi_{j-1+m}[\int_0^\infty tdF(t)]^{-1}\int_0^\infty dG(t)\int_0^t\binom{j+m}{m}e^{-j\mu x}(1-e^{-\mu x})^m dx$$

for $j\geq 1$. Taking generating functions and simplifying as before through the normalising condition we find

$$q(z)=[\int_0^\infty tdG(t)]^{-1}\int_0^\infty dG(t)\int_0^t w[1+(z-1)e^{-\mu x}]dx$$
$$= 1+A\sum_{1}^{\infty}(z-1)^m h_{m-1}/m, \quad (9)$$

where A is the ratio of the arrival rate to the service rate. This gives that the mean traffic $M = q'(1) = A$, so that $M = -[\mu\psi'(0)]^{-1}$. Differentiation of (9) gives the relation

$$q^{(n)}(1)/n! = Ah_{n-1}/n, \quad n\geq 1, \quad (10)$$

for the binomial moments of the queue distribution. Also

$$q^{(n)}(1) = A\pi^{(n-1)}(1), \quad n\geq 1, \quad (11)$$

relating the moments of the q- and π-distributions. Finally, (9) gives

$$q_j = A \sum_{m=j}^{\infty}(-1)^{m-j} \binom{m}{j}h_{m-1}/m, \quad j\geq 1,$$
$$q_o = 1 + A \sum_{m=1}^{\infty}(-1)^m h_{m-1}/m.$$

Although the expressions for the π_j and q_j are not in finite form they lead to rapid convergents and so are convenient for calculation. Thus

$$h_n = h_{n-1}\phi(n\mu)/[1-\phi(n\mu)] \sim h_{n-1}\phi(n\mu)$$

for n large, and $\phi(n\mu)$ decreases rapidly. For example $\phi(n\mu) \sim 1/n$ if ϕ represents a negative exponential distribution, and the h_n decrease at a rate comparable to the terms of an exponential series.

A number of other quantities of practical interest are easily derived in terms of our formalism. Thus the generating function $Q(z)$ for the steady-state probability of j or more channels being occupied is given by

$$Q(z)=q(z)+[1-q(z))]/(1-z)=1+A[1+\sum_{m=1}^{\infty}(z-1)^m\{h_{m-1}/m+h_m/(m+1)\}]$$

Thus the probabilities $\{Q_j;\ j\geq 0\}$ are derived from the q_j simply by replacing h_{m-1}/m by $h_{m-1}/m + h_m/(m+1)$.

The explicit forms for the π- and q- distributions show that knowledge of either prescribes $\psi(n\mu)$ for each $n\geq 1$ (also $\psi'(0)$ in the case of q). Because of the monotone properties of the derivatives $\psi^{(m)}(s)$ for real s this means that the shape of the function $\psi(s)$ is fairly tightly prescribed, so that for practical purposes there is only one distribution function G for the times separating successive calls that can result in a given observed traffic.

4. BASIC CHARACTERISTICS OF TRAFFIC

A number of conclusions concerning traffic can be drawn from the results of the last section without use being made of any special assumption as to the nature of the inter-event time distribution G. Denote by m, v the mean and variance of the distribution G. Since $m = -\psi'(0)$ and $M = -[\mu\psi'(0)]^{-1}$, the mean traffic M is inversely proportional to the expected time between calls, that is, M is proportional to the intensity of the stream. We are so accustomed to thinking of channel occupancy as the way of measuring traffic that this may appear tautological. However, a theorem is implicit here: the mean number of channels occupied in an $G/M/\infty$ queueing system in the steady-state is equal to the ratio A of the mean arrival rate to the mean service rate. In other words, we have the result M = A of the last section.

Let us now turn our attention to second moments. Since $q''(1) = Mh_1$, the variance $V = q''(1)+q'(1)-[q'(1)]^2$ of the traffic is given by

$$V = M[1-M+h_1] = M[1-M+\psi(\mu)/(1-\psi(\mu))] \qquad (12)$$

which is a version of the well-known Molina-Nyquist relation

$$V = M[1-M+A/(N+1+M-A)],$$

giving the variance V of overflow traffic arising from a Poisson traffic of A erlangs offered to N channels (see Riordan's appendix to Ref. 1). In that case $\phi(s)=\lambda/(\lambda+s)$ and $k_n(\mu) = n!(A+n+1)/[A^n(A+1)]$, $n\geq 0$, where $A = \lambda/\mu$. Thus for the overflow

$$\psi(\mu)/(1-\psi(\mu)) = \Big[\sum_{r=0}^{N-1}\binom{N-1}{r}k_r(\mu)\Big]\Big/\Big[\sum_{r=1}^{N}\binom{N-1}{r-1}k_r(\mu)\Big]$$

$$= \Big[A\sum_{m=0}^{N}A^m/m!\Big]\Big/\Big[(N+1)\sum_{m=0}^{N}A^m/m! - A\sum_{r=0}^{N-1}A^m/m!\Big]$$

$$= A/(N+1+M-A)$$

by Erlang's loss formula, agreeing with the Molina-Nyquist formula. Our formula (12) is more general, holding for an arbitrary initial renewal stream of calls or the resultant traffic from such a stream at any stage of a tandem sequence of overflows. Similar formulae can be derived readily relating to higher moments of such traffic. (12) can also be written in the form $V=M[1-M+M^*]$, where M^* is the mean of the π-distribution.

Relation (12) may be given an intuitive interpretation. Roughly speaking, for a given value of M (or equivalently of m or $\psi'(0)$), the expression $\psi(\mu)/(1-\psi(\mu))$ will be large when ψ'' is large, since $\psi(0) = 1$. As $\psi''(0) = m^2+v$ and ψ'' is monotone decreasing for a positive argument, this in turn corresponds to large v. Thus, roughly speaking, for a fixed mean traffic the variance V of the traffic is large when the variance v of the inter-event times in the stream of arriving calls is large.

5. CARRIED TRAFFIC IN TERMS OF G

In the spirit of Section 3, imagine a stream of calls characterised by distribution function G offered to N channels in parallel and let $\{\pi_j;\ 0\leq j\leq N\}$, $\{q_j;\ 0\leq j\leq N\}$ represent respectively the imbedded chain and continuous time distributions in the steady-state for the number of busy channels induced. Then for $0\leq j\leq N$

$$\pi_j = \sum_{m=j-1}^{N-1}\pi_m\int_0^\infty\binom{m+1}{j}e^{-j\mu x}(1-e^{-\mu x})^{m+1-j}dG(x)$$

$$+ \pi_N\int_0^\infty\binom{N}{j}e^{-j\mu x}(1-e^{-\mu x})^{N-j}dG(x).$$

In terms of the generating function $\pi(z)=\sum_{j=0}^{N}\pi_j z^j$ these relations may be expressed as

$$\pi(z)=\int_0^\infty w[1+(z-1)e^{-\mu x}]dG(x)-\pi_N\int_0^\infty(z-1)e^{-\mu x}[1+(z-1)e^{-\mu x}]^N dG(x)$$

where as before $w(z)=z\pi(z)$ and we have used the normalising condition $\pi(1)=1$. Analogously to (7) we have

$$\pi^{(n)}(1)/n!=[\pi^{(n-1)}(1)/(n-1)!]\psi(n\mu)/(1-\psi(n\mu))$$

$$- \pi_N\binom{N}{n-1}\psi(n\mu)/(1-\psi(n\mu)),$$

whence

$$\pi^{(n)}(1)/n! = h_n\Big[1-\pi_N\sum_{m=0}^{n-1}\binom{N}{m}h_m^{-1}\Big].$$

Since $\pi^{(N)}(1)/N! = \pi_N$, we have for the probability of loss

$$\pi_N = \Big[\sum_{m=0}^{N}\binom{N}{m}h_m^{-1}\Big]^{-1}. \qquad (13)$$

In the case where G is negative exponential with parameter λ this reduces to

$$\pi_N = (A^N/N!)/\Big[\sum_{r=0}^{N}A^r/r!\Big],$$

with $A = \lambda/\mu$, which is just the Erlang loss formula used in the previous section. An explicit expression for $\pi(z)$ follows from the Taylor series

$$\pi(z) = \sum_{n=0}^{N}(z-1)^n\,\pi^{(n)}(1)/n!$$

The q-distribution may be found from the relations

$$q_j=\sum_{m=j-1}^{N-1}\pi_m\Big[\int_0^\infty t\,dG(t)\Big]^{-1}\int_0^\infty dG(t)\int_0^t\binom{m+1}{j}e^{-j\mu x}(1-e^{-\mu x})^{m+1-j}dx$$

$$+ \pi_N\Big[\int_0^\infty t\,dG(t)\Big]^{-1}\int_0^\infty dG(t)\int_0^t\binom{N}{j}e^{-j\mu x}(1-e^{-\mu x})^{N-j}dx$$

which holds for $0 < j \leq N$. This leads to the generating function

$$q(z)=\Big[\int_0^\infty t\,dG(t)\Big]^{-1}\int_0^\infty dG(t)\int_0^t w[1+(z-1)e^{-\mu x}]dx$$

$$-\pi_N\Big[\int_0^\infty t\,dG(t)\Big]^{-1}\int_0^\infty dG(t)\int_0^t(z-1)e^{-\mu x}[1+(z-1)e^{-\mu x}]^N dx$$

$$= 1 + A\sum_{n=1}^{N}(z-1)^n[\pi^{(n)}(1)/n!][1-\psi(n\mu)]/(n\psi(n\mu)),$$

where $A = [\mu\int_0^\infty t\,dG(t)]^{-1}$ is the mean offered traffic. We deduce in particular that

$$q^{(n)}(1) = A[\pi^{(n-1)}(1)-\pi_N(N)_{n-1}],\quad n>0. \qquad (14)$$

The mean carried traffic $M = q'(1)$ is thus given by

$$M = A[1-\pi_N] \qquad (15)$$

as we would expect. Equations (13) and (15) give a general version of the standard Erlang loss formula.

The formulae of this section give rise to various "carried traffic" analogues of the ordinary traffic formulae of the last section. (14) and (15) are of course analogues to (11) and the relation M = A of Section 3. With second moments we have

$$V = q''(1)+q'(1)-[q'(1)]^2$$
$$= A[h_1-\pi_N(h_1+N)]+M-M^2$$
$$= M[1-M+h_1-N\pi_N/(1-\pi_N)].$$

We note that

$$N\pi_N/(1-\pi_N)=N/[\sum_{m=1}^{N} \binom{N}{m}h_m^{-1}]=N/[Nh_1^{-1}+\sum_{m=2}^{N} \binom{N}{m}h_m^{-1}] \leq h_1,$$

with equality only for N=1. Thus $V \geq M[1-M]$, with equality only for N=1. It can be seen that $N\pi_N/(1-\pi_N)$ decreases monotonically with increasing N.

6. OVERFLOW TRAFFIC FORMULAE

We now continue the results of Section 4 with the additional requirement that the distribution G arise specifically as overflow from a set of N channels of a stream characterised by some distribution function F. We begin with the basic relation (5), which because of $k_r(0)=\delta_r(0)$ leads to

$$\psi'(0)=\sum_{r=0}^{N-1} \binom{N-1}{r}k_r'(0)-\sum_{r=0}^{N} \binom{N}{r}k_r'(0) = -\sum_{r=1}^{N} \binom{N-1}{r-1}k_r'(0).$$

Likewise $k_r(s) = \prod_{j=1}^{r} [1-\phi(s+(j-1)\mu)]/\phi(s+j\mu)$ with $\phi(0)=1$ provides

$$k_r'(0) = -[\phi'(0)/\phi(\mu)] \prod_{j=2}^{r} [1-\phi((j-1)\mu)]/\phi(j\mu), \quad r>0,$$

where again we interpret the empty product as unity. Hence

$$k_r'(0) = -k_{r-1}(\mu)\phi'(0)/\phi(\mu)$$
$$\psi'(0) =[\phi'(0)/\phi(\mu)] \sum_{r=0}^{N-1} \binom{N-1}{r}k_r(\mu). \qquad (16)$$

For each r $k_r(\mu)/\phi(\mu)$ is, for fixed $\phi'(0)$(and $\phi(0)=1$) a monotone decreasing function of ϕ. Arguing as for ψ in Section 4, we see that for fixed $\phi'(0)$, the RHS of (16) is small when $\phi''(0)$ is large, that is, when the distribution given by F has a large variance. Hence the mean traffic $M = -[\mu\psi'(0)]^{-1}$ is increased, for a fixed mean traffic in the F-stream, by increases in the variance of inter-event times in that stream. This is as one might expect. It is also apparent that in general the mean overflow traffic depends on higher characteristics of the input stream than just its mean and variance. This makes it implausible that simple relations analogous to those of Olsson (Ref. 13) will hold except perhaps for specific forms of input stream which can be completely characterised by a small number of parameters. Signal amongst such possibilities is the Poisson stream, described by a single parameter. We look at this possibility more closely in Section 7.

The special properties of the k's allow for a convenient representation of higher order overflows in terms of the distribution function of the initially offered traffic stream and the numbers of channels at each stage. Thus suppose an input stream F overflows from N_1 channels and the overflow itself overflows from N_2 secondary channels to give traffic characterised by $\psi_{(2)}(s)$. Suppose $k_{(2)r}$ stands to $\psi_{(2)}$ as does k_r to ψ. Directly from (4)

$$k_{r+1}(s+(m-1)\mu)=k_r(s+m\mu)[1-\phi(s+(m-1)\mu)]/\phi(s+m\mu),$$

so that

$$\sum_{r=1}^{N_1} \binom{N_1-1}{r-1}k_r(s+(j-1)\mu)=[\sum_{r=0}^{N_1-1} \binom{N_1-1}{r}k_r(s+j\mu)] \times$$
$$[1-\phi(s+(j-1)\mu)]/\phi(s+j\mu).$$

Thus

$$[1-\psi(s+(j-1)\mu)]/\psi(s+j\mu) =$$

$$\frac{\sum_{r=1}^{N_1} \binom{N_1-1}{r-1}k_r(s+(j-1)\mu)}{\sum_{r=0}^{N_1} \binom{N_1}{r}k_r(s+(j-1)\mu)} \div \frac{\sum_{r=0}^{N_1-1} \binom{N_1-1}{r}k_r(s+j\mu)}{\sum_{r=0}^{N_1} \binom{N_1}{r}k_r(s+j\mu)} =$$

$$\frac{1-\phi(s+(j-1)\mu)}{\phi(s+j\mu)} \times \frac{\sum_{r=0}^{N_1} \binom{N_1}{r}k_r(s+j\mu)}{\sum_{r=0}^{N_1} \binom{N_1}{r}k_r(s+(j-1)\mu)}$$

and

$$k_{(2)m}(s) = \prod_{j=1}^{m} [1-\psi(s+(j-1)\mu)]/\psi(s+j\mu)$$
$$= k_m(s)[\sum_{r=0}^{N_1} k_r(s+m\mu)\binom{N_1}{r}]/[\sum_{r=0}^{N_1} \binom{N_1}{r}k_r(s)]$$
$$=[\sum_{r=0}^{N_1} \binom{N_1}{r}k_{m+r}(s)]/[\sum_{r=0}^{N_1} \binom{N_1}{r} k_r(s)].$$

Therefore

$$\psi_{(2)}(s) =[\sum_{m=0}^{N_2-1} \binom{N_2-1}{m}k_{(2)m}(s)]/[\sum_{m=0}^{N_2} \binom{N_2}{m}k_{(2)m}(s)]$$
$$= \frac{\sum_{m=0}^{N_2-1} \sum_{r=0}^{N_1} \binom{N_2-1}{m}\binom{N_1}{r}k_{m+r}(s)}{\sum_{m=0}^{N_2} \sum_{r=0}^{N_1} \binom{N_2}{m}\binom{N_1}{r}k_{m+r}(s)},$$

a strikingly simple generalisation of (5). We derive

$$\psi'_{(2)}(0)= [\phi'(0)/\phi(\mu)] \sum_{m=0}^{N_2-1} \sum_{r=0}^{N_1} \binom{N_2-1}{m}\binom{N_1}{r}k_{m+r}(\mu),$$

the second overflow version of (16). This formula is convenient for determining the mean traffic

$$M = -[\mu\psi'_{(2)}(0)]^{-1}$$

after the second overflow. The corresponding formula for the variance V is

$$V = M[1-M+\psi_{(2)}(\mu)/(1-\psi_{(2)}(\mu))].$$

We obtain the term in ψ from

$$\frac{\psi_{(2)}(\mu)}{1-\psi_{(2)}(\mu)} = \frac{\sum_{m=0}^{N_2-1} \sum_{r=0}^{N_1} \binom{N_2-1}{m}\binom{N_1}{r}k_{m+r}(\mu)}{\sum_{m=1}^{N_2} \sum_{r=0}^{N_1} \binom{N_2-1}{m-1}\binom{N_1}{r}k_{m+r}(\mu)}$$

These various formulae extend to n-th iterated overflow in an obvious way.

7. GENERAL RESULTS

Although our formulae are designed for specific calculations, their generality enables us to consider a number of broader questions, especially those of an existence type, which would be much less accessible with other approaches. We consider four such examples in this section.

7.1 NEGATIVE BINOMIAL DISTRIBUTION

Although its applicability is known to be only approximate the negative binomial distribution has often been used to describe overflow traffic. The appeal of this distribution is multiple, residing in its simplicity, the fact that it gives $V > M$, and in that its use can be justified by a plausible heuristic involving birth coefficients. In view of these strengths, it would be interesting to know the nature of a traffic stream (perhaps itself an overflow) whose overflow was exactly negative binomial. Unfortunately we can see as follows that there is no such stream:

From (10), we have that for any traffic

$$[q^{(n+1)}(1)/(n+1)!]/[q^{(n)}(1)/n!]$$

$$= (h_n/h_{n-1})n/(n+1) = [\phi(n\mu)/(1-\phi(n\mu))]n/(n+1) \to 0 \text{ as } n\to\infty.$$

For the negative binomial distribution, we have

$$q(z) = [(1-q)/(1-qz)]^m$$

for some m, so that

$$[q^{(n+1)}(1)/(n+1)!]/[q^{(n)}(1)/n!]=[(m+n)/(n+1)]q/(1-q)$$

which does not tend to 0 as n tends to ∞, giving the required contradiction. Our argument indicates that the reason for this failure is that the negative binomial distribution contains too much probability in its tail. This suggests that its practical use is fail-safe in that it may be expected to over-estimate the probability of high channel occupancy.

7.2 ASYMPTOTIC BEHAVIOUR

It might be supposed that although successive overflow streams become less intense, repeated overflowing results in some stable asymptotic functional form for the structure of the overflow stream, that is, a functional form which occurs with both the offered stream and the overflow. Such a stream would be of considerable interest in modelling compound teletraffic systems.

Most simply we can try allowing for the intensity reduction by positing a simple time dilatation, so that $G(x) = F(ax)$ with $0<a<1$. Thus

$$\psi(s) = \phi(s/a). \qquad (17)$$

Consider the simplest case N = 1, that is, of overflow from a single channel. Then

$$\phi(s/a)=\psi(s)=[1+k_1(s)]^{-1}=\phi(s+\mu)/[\phi(s+\mu)+1-\phi(s)].$$

Thus

$$\phi(s/a)/[1-\phi(s/a)]=\phi(s+\mu)/[1-\phi(s)],$$

and as ϕ is strictly monotone decreasing for positive arguments, we have

$$\phi(s+\mu)/[1-\phi(s+\mu)]<\phi(s/a)/[1-\phi(s/a))]< \phi(s)/[1-\phi(s)]$$

for s positive. But the strict monotonicity of $x/(1-x)$ and ϕ now leads to

$$s+\mu > s/a > s,$$

which clearly cannot hold identically for s > 0, a contradiction. A similar objection occurs if we try to allow for a superimposed shift of origin.

More generally, with an arbitrary N, we have asymptotically for large positive s

$$\psi(s) \sim k_{N-1}(s)/k_N(s) \sim \phi(s+N\mu),$$

so that by (17)

$$\phi(s/a)/\phi(s+N\mu) \to 1 \text{ as } s \to +\infty .$$

This can be seen to be incompatible with the possible behaviour of G(t) at the origin.

Essentially equivalent objections arise if we allow slowing of the service rate at the second stage to $a\mu$ but require agreement only of the channel occupancies' measure of the traffics given by F and G. Our formulae (10) for the q-distribution then give $\phi'(0)=a\psi'(0)$ and $\phi(n\mu) =\psi(na\mu)$ for all $n\geq 1$, leading to the same difficulties. This points to a continuing structural change in the form of traffic with successive overflows, as might have been foreseen by our earlier observation that the variance and higher characteristics of offered traffic affect the mean of the overflow.

7.3 V > M

While it is well known that the traffic resulting from the overflow of an offered Poisson stream has V > M, we may note that this property does not hold invariably for all forms of input. Arguing intuitively, we expect that when the intensity of the input stream is sufficiently high overflowing will not appreciably increase the variance of the offered traffic and the overflow stream will appear much like the input stream. Since for a given value M, V increases with the variance of the distribution G, we would expect that V < M might occur for all input streams of sufficiently high intensity and low variance for the distribution F. This heuristic can be substantiated analytically.

For traffic of sufficiently high intensity $\phi(j\mu) \sim 1$ for $j = 1, .., N+1$, so that we have (to second order)

$$M \sim -\phi(\mu)/[\mu\phi'(0)[1+(N-1)k_1(\mu)]],$$

$$\psi(\mu)/[1-\psi(\mu)] \sim [1+(N-1)k_1(\mu)]/[k_1(\mu)+(N-1)k_2(\mu)].$$

In the extreme case of least variance when successive arrival instants in the offered stream have constant separation d, $\phi(s) = \exp(-sd)$. For heavy traffic $\mu d \sim 0$ and the expressions above become (to second order)

$$M \sim (1-N\mu d)/(\mu d),$$

$$\psi(\mu)/(1-\psi(\mu)) \sim [1-(N+1/2)\mu d]/(\mu d),$$

so that by (12) $V \sim M/2$ to first order. Thus V < M is theoretically possible.

7.4 A CONJECTURE OF OLSSON

Although it is not our intention in this paper to treat the very difficult problem of pooled traffic, we consider one simple situation involving pooled traffic as our results show that a putative approximate formula of Olsson holds exactly in a simple case. Olsson's result (which is given with a typographical error in Ref. 13) states that if n streams of traffic characterised by means and variances M_i, V_i $(i = 1, 2,..,n)$ are jointly offered to a set of channels, then the means of the overflow traffics are in the proportion of the quantities $V_i + M_i^2/V_i$.

Consider the case of n Poisson streams, characterised by parameters λ_i, feeding into a common set of channels. By the lack of memory property of a Poisson stream, each call of the aggregate stream (itself Poisson with parameter $\lambda = \Sigma\lambda_i$) has the same probability $p_i = \lambda_i/\lambda$ of having arisen from stream i. Since the overflow does not distinguish between the origins of the calls, the same property will hold for each successive overflow. Hence in particular the mean overflow intensities and hence the mean traffics will be in the proportion of the original input intensities λ_i. This agrees with Olsson's formula, which for Poisson streams reduces to overflow means being proportional to $2\lambda_i$.

Less trivially, suppose that after one or more overflows the distribution function for the aggregate stream is G(x). Then the inter-event times for the overflows from stream i will be

$$p_i G(x)+p_i(1-p_i)G^{(2)}(x)+p_i(1-p_i)^2G^{(3)}(x)+..,$$

where $G^{(m)}(x)$ is the m-th convolution of G(x). The corresponding Laplace-Stieltjes transform $\psi_i(s)$ can be expressed in terms of the transform $\psi(s)$ for the aggregate stream as

$$\psi_i(s) = p_i\psi(s)\sum_{m=0}^{\infty} [(1-p_i)\psi(s)]^m = p_i\psi(s)/[1-(1-p_i)\psi(s)].$$

This is readily seen to give $\psi_i'(0)=p_i^{-1}\psi'(0)$, so that the corresponding mean traffic

$$M_i = -[\mu\psi_i'(0)]^{-1} = -p_i/[\mu\psi'(0)],$$

and the mean traffics are as the λ_i, as we saw above. However, our formula for $\psi_i(s)$ enables us to derive corresponding results for higher moments. Thus for the variances we have

$$V_i = M_i[1-M_i+\psi_i(\mu)/(1-\psi_i(\mu))]= M_i[1-M_i+p_i\psi(\mu)/(1-\psi(\mu))]$$

which evaluates to give for the first overflow

$$V_i = M_i[1-M_i+\lambda_i/(N+M+1-\lambda)],$$

where $M = \Sigma M_i$. This formula has been found by alternative means by Wilson (Ref. 12) and our method extends to a number of other cases. Thus if the holding times are different for the different input streams (reflecting, perhaps, differential cost factors for calls for different origins) the mean traffics M_i are proportional to λ_i/μ_i and the V_i are given by

$$V_i = M_i[1-M_i+p_i\psi(\mu_i)/(1-\psi(\mu_i))].$$

8. SUMMARY

By measuring teletraffic streams by the times separating consecutive calls rather than by the number of busy channels induced in an infinite set by the calls, it is possible to analyse stochastic behaviour in compound teletraffic networks piecemeal, considering traffic at each stage separately without reference to previous stages. The basis of the technique is the explicit characterisation of an overflow stream in terms of the offered stream and the number of channels to which the traffic is offered. Actual channel occupancy probabilities are then calculated from the specification of the relevant stream of calls using methods of queueing theory.

Generalisations were derived for the standard expressions for the mean and variance of the overflow traffic from a set of channels. The results derived hold for arbitrary non-random renewal streams of calls and after an arbitrary number of sequential overflows.

The generality and simplicity of the approach were illustrated by consideration of four general questions: the possibility of exactly negative binomial overflow traffic, the asymptotic behaviour of overflow streams, the possibility of the mean overflow traffic exceeding the variance, and Olsson's conjecture.

9. REFERENCES

1. Wilkinson, R.I., "Theories for Toll Traffic Engineering in the U.S.A.", Bell System Technical Journal, Vol. 35, No. 2, pp. 421-514, 1956, (appendix by J. Riordan).

2. Neovius, G., "Artificial Traffic Trials using digital computers", Ericsson Technics, Vol. 11, No. 2, pp. 279-291, 1955.

3. Le Gall, P., "Le trafic de débordement", Annales des télécommunications, Vol. 16, Nos. 9-10, pp.226-238, 1961.

4. Palm, C., "Intensitatsschwankungen im Fernsprechverkekr", Ericsson Technics, Vol. 44, pp. 1-189, 1943.

5. Takács, L., "On secondary stochastic processes generated by recurrent processes", Acta Math. Acad. Sci. Hung., Vol. 7, No. 1, pp. 17-30, 1956.

6. Kosten, L., "Sur la probabilité d'encombrement dans les multiplages gradués", Annales des P.T.T., pp.1002-1019, 1937.

7. Kosten, L., "Uber Sperrungswahrscheinlichkeiten bei Staffelschaltungen", E.N.T., Vol. 14, p.5, Jan. 1937.

8. Takács, L., "Introduction to the Theory of Queues", Oxford University Press, New York, 1962.

9. Cohen, J.W., "The full availability group of trunks with an arbitrary distribution of the inter-arrival times and the negative exponential holding time distribution", Simon Stevin Wis-en Natuurkundig Tijdschrift, Vol. 26, pp. 169-181, 1957.

10. Brockmeyer, E., "Det simple overflowproblem i telefontra ficteorien", Teleteknik, Vol. 5, No. 4, pp. 361-374, 1954.

11. Descloux, A., "On overflow processes of trunk groups with Poisson inputs and exponential service times", Bell System Technical Journal, Vol. 42, No. 1, pp. 383-397, 1963.

12. Potts, R.B., Berry, L.T.M., Bruyn, S.J., and Wilson, K.G., "University of Adelaide-Australian Post Office, Adelaide Telephone Network Study, Second Progress Report", June 1975.

13. Pratt, C.W., "Report on Overseas Visit", Australian Post Office Headquarters, Melbourne, 1967.

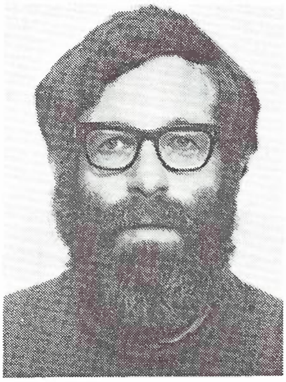
Discussion

P. LE GALL, France : *Section 7.4 deals with the combination of several infinite overflow traffics. How is it possible that the variance of the ith overflow traffic depends on the total overflow traffic.*

Why can you write: "Since the overflow does not distinguish between the origins of the calls".

C.E.M. PEARCE and R.M. POTTER, Australia : *Considering the first part of the question; overflows can occur only when all the trunks at the first stage are occupied, and this event will depend on the total offered traffic. The variance of the i^{th} overflow traffic will thus depend on the total offered traffic. This latter traffic is most conveniently brought in to the formulae through the total overflow traffic.*

Turning to the second part of the question, I think our wording was poorly chosen. What was meant was that by virtue of the lack of memory property of Poisson streams, any knowledge of trunk occupancy tells us nothing about the origin of a given call in the overflow.



BIOGRAPHY

CHARLES PEARCE received his early training in New Zealand in Mathematics and Physics. He subsequently came to Australia where he completed a Ph.D. in the area of Stochastic Processes at Canberra in 1965. After a period lecturing in Britain he returned to Australia and joined the University of Adelaide in 1968 as a senior lecturer in Mathematics.

BIOGRAPHY

RONDA POTTER graduated B.Sc. with first class honours in Mathematics at Adelaide University in 1966, and completed her M.Sc. in 1968. She is a senior tutor in the Department of Applied Mathematics and has interests in Operations Research generally. She is currently working towards a Ph.D. in the area of mathematical models for teletraffic under the supervision of Dr. Charles Pearce.

The Accuracy of Overflow Traffic Models

A. H. FREEMAN

Telecom Australia, Sydney, Australia

ABSTRACT

This paper compares the equivalent random model for overflow traffic with two models based on the interrupted poisson process. One of these models is considerably more difficult to compute than the E.R. model but its accuracy is much greater and it is useful as a reference for comparing different models. The other is of comparable accuracy to the E.R. model and in some applications is more easily computed.

1. INTRODUCTION

The most widely accepted procedures for calculations with overflow traffic is Wilkinsons Equivalent Random (ER) model, and variations of it (Ref. 1). Nearly 20 years of practical experience have firmly established its adequacy for practical applications yet there is suprisingly little numerical data on its accuracy. This paper reports the results of an investigation of an alternative model using the I.P.P. process which throws some light on the nature of the ER approximation.

2. PRINCIPLES OF INVESTIGATION

A typical overflow configuration is shown in figure 1. There are "r" distinct sources, each overflowing into a common route, and the usual problem is to determine the size of the common route to satisfy some service criterion.

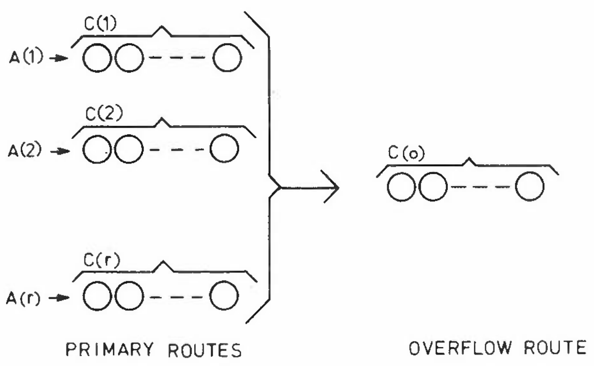


FIGURE 1 Typical Overflow Route

In principle, this problem can be solved by state equations, but in practice the number of states is unmanageably large, for figure 1 the number being $(C(1)+1)(C(2)+1) - - - (C(r)+1(C(o)+1)$.

The E.R. model replaces the r sources with a single source of A(e) erlangs offered to C(e) circuits, such that the mean and variance of the overflow from this single source is identical to that of the r actual sources, giving the model of figure 2. This model has only $C(e)+C(o)+1$ states and its solution is very simple.

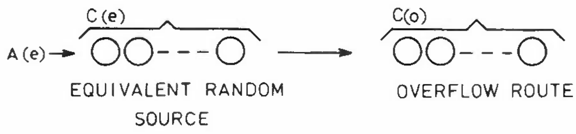
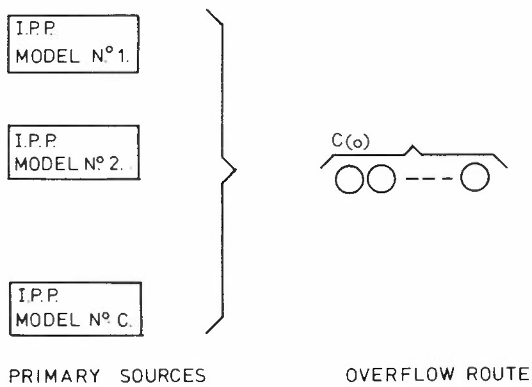


FIGURE 2 Equivalent Random Model

Kuczura (Ref.2) describes the Interrupted Poisson Process (I.P.P.) which gives a remarkably close approximation to the overflow from a single route. This is a poisson process modulated by a switch which is alternately turned on and off, with the on and off periods being independent negative exponential variables. It is defined by three parameters, ℓ, w and g where:-

ℓ is the intensity of the poisson process.
1/w is the mean on time of the switch.
1/g is the mean off time of the switch.

This model has two states and the three parameters can be used to match the first three moments of a distribution. If each of the sources of figure 1 is replaced by an I.P.P. approximation one gets the model of figure 3, in which the number of states is $2^{(r)}.(C(o)+1)$, considerably less than for the actual configuration, but greater than the E.R. model of figure 2. It will be shown that the model of figure 3 is generally more accurate than the E.R. model, and because it takes account of the actual structure of the sources it allows some conclusions to be drawn about the accuracy of the E.R. model.



NOTE Each ER model replaces one of the sources in fig. 1

FIGURE 3 I.P.P. Accurate Model

3. POTENTIAL ACCURACY

Ref. (2) gives some comparisons of the probability distributions of the overflow from a full availability route with those of an I.P.P. model matching the first three moments. The differences shown are small and suggest that the accuracy of the model may be sufficient for the purpose in mind.

A more direct assessment of the potential accuracy was made by comparison with some cases for which exact solutions are available. Use was made of results published by Kibble (Ref.3), giving exact solution for a number of cases with two routes overflowing into a common route, as well as the approximation using the E.R. model. For each case a solution using the I.P.P. approximation of Fig.3 was obtained and compared. Out of 97 cases, the I.P.P. model was more accurate in all but 2 cases. The I.P.P. model was usually from 5 to 20 times as accurate as the E.R. model, and in the cases where the difference was small or where E.R. was better, both were very close approximations.

Further evidence on the accuracy of the I.P.P. model arose as a consequence of the actual study.

4. DETAILS OF INVESTIGATION

For the purposes of this investigation, 9 different combinations of high usage routes were chosen, each of which generated an overflow traffic of 10E with a variance of 20. For each of these the E.R. model uses the same model but the I.P.P. model is different for each case, and any differences in the results include the effect of differences in the make-up of the overflow traffic which are ignored in the E.R. method.

The choice of the overflow traffic mean and variance was a compromise between the desirability of a small value for which computing would be more economical and of using values typical of actual routes. Although rather low, traffic of this order is found on some high usage routes in Australia. The combinations were chosen for convenience of calculation with the possibility of detecting a significant pattern also kept in mind. Table 1 lists these combinations.

Case No.	Composition
1	Single Source M=10, V=20 (E.R. Model)
2	Two Sources each M=5, V=10
3	Two Sources (one M=5, V=5 (one M=5, V=15
4	Three Sources (one M=5, V=10 (two M=2.5, V=5
5	Four Sources each M=2.5, V=5
6	Eight Sources each M=1.25, V=2.5
7	16 Sources each M=5/8, V=1.25
8	32 Sources each M=5/16, V=5/8
9	64 Sources each M=5/32, V=5/16

TABLE 1

For each of the components of these combinations the following were calculated.

(1) The offered traffic (A) and the number of circuits (C) from which overflow of the specified mean and variance would be generated.

(2) The first seven factorial moments of the overflow traffic component.

(3) The I.P.P. parameters to generate traffic with the same first three factorial moments.

Case No.	FACTORIAL MOMENTS $F3/10^3$	$F4/10^4$	$F5/10^5$	$F6/10^6$	$F7/10^7$	
1	1.2988	1.6214	2.1187	2.8770	4.0376	
2	1.3042	1.6441	2.1825	3.0287	4.3689	
3	1.3201	1.7101	2.3652	3.4595	4.9329	
4	1.3064	1.6537	2.2108	3.0995	4.5361	
5	1.3085	1.6633	2.2392	3.1703	4.6949	
6	1.3117	1.6774	2.2817	3.2799	4.9563	
7	1.3137	1.6867	2.3102	3.3548	5.1385	
8	1.3150	1.6926	2.3286	3.4033	5.2582	
9	1.3158	1.6962	2.3400	3.4338	5.3340	
1(IPP)	1.2988	1.6199	2.1107	2.8505	3.9658	(2)
NEG. BIN	1.32	1.716	2.4024	3.6304	5.7658	(3)

Note 1. F1=10, F2=110 for all cases.
2. Moments of IPP fitting case 1.
3. Moments of Negative binomial with F1=10, F2=110.

TABLE II

The factorial moments of the combinations were then calculated from those of the components, and are listed in Table II. Also for comparison is given the moments of a negative binomial distribution of the same mean and variance. Examination of the table suggests the following hypotheses.

(1) For any combination of overflows from full availability routes the third and higher moments are higher than those of the matching ER model.

(2) If all the components have the same ratio of variance to mean the moments approach those of a negative binomial, as the number of components increases.

(3) If the components have different ratios of variance to mean, the third and higher moments increase more rapidly than if the ratios are equal. Cases have been found where they exceed those of the corresponding negative binomial.

Although it has not been possible to prove these hypotheses rigorously, a considerable amount of numerical work supports them.

Using the parameters referred to above the grade of service for overflow routes of 1 to 24 circuits was calculated using the ER model and the various I.P.P. models. These results are listed in Table III together with those for two models both generating negative binomial traffic.

CCTS	Case No. 1	1A	2	3	4	5	6
1	.91667	.91667	.91477	.91191	.91411	.91350	.91280
2	.83527	.83527	.83165	.82619	.83040	.82922	.82786
3	.75607	.75607	.75095	.74330		.74749	.74552
4	.67936	.67935	.67300	.66370	.67083	.66866	
5	.60547	.60543	.59814	.58791		.59312	
6	.53473	.53466	.52675	.51635	.52401	.52127	.51803
7	.46751	.46740	.45921	.44943		.45350	
8	.40420	.40403	.39591	.38745	.39309	.39020	
9	.34516	.34490	.33721	.33061		.33172	.32840
10	.29076	.29042	.28342	.27902	.28093	.27838	
11	.24120	.24084	.23479	.23271		.23037	
12	.19700	.19644	.19151	.19162	.18969	.18782	.18554
13	.15802	.15734	.15362	.15563		.15072	
14	.12437	.12359	.12105	.12456		.11896	
15	.09591	.09506	.09362	.09813		.09227	.09145
16	.07240	.07150	.07099	.07603	.07066	.07030	
17	.05344	.05254	.05274	.05789		.05257	
18	.03854	.03767	.03836	.04328		.03858	.03881
19	.02714	.02633	.02731	.03174		.02777	
20	.01866	.01794	.01902	.02283	.01937	.01961	
21	.01252	.01190	.01296	.01609		.01356	.01413
22	.00820	.00769	.00863	.01111		.00919	
23	.00523	.00484	.00562	.00751		.00609	
24	.00327	.00296	.00358	.00497	.00385		.00444
Note	1	2	3	3	3	3	3

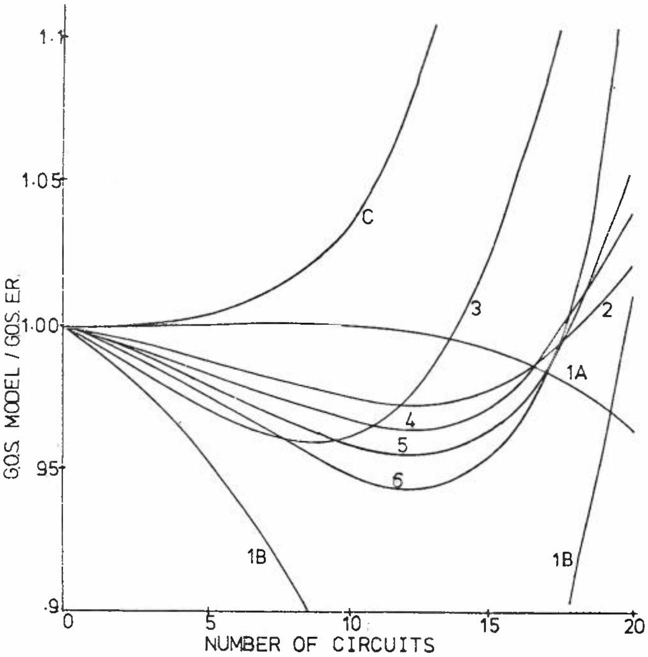
Notes 1. G.O.S. using E.R. model
2. G.O.S. using I.P.P. approximation to 1
3. G.O.S. using model of fig. 3.

TABLE III

5. ANALYSIS OF RESULTS

In the case of a single source the E.R. model (model 1) is an exact solution and the I.P.P. (model 1A) an approximation. Therefore comparison of these will give an indication of the accuracy of the latter. For the other combinations only the I.P.P. approximation are available, but they differ from model 1A by considerably more than model 1A differs from the exact solution of model 1. Therefore the differences are a good indication of the differences which would arise in an exact solution.

In order to illustrate this better figure 4 shows the difference between the grades of service of each model and model 1. Certain trends are clearly visible.



(1) For all the cases with more than one source the congestion is lower than for a single source for moderate numbers of circuits, but is higher for large numbers of circuits. Therefore the ER model always overestimates congestion for poor grades of service and underestimates for good grades of service.

(2) The difference curves do not correlate clearly with the higher moments, or with any other parameter which would lead to a convenient formula of greater accuracy than the ER model.

(3) Although the magnitude of the differences is small enough to be ignored in practical calculations, they may be significant in theoretical investigations of for example the difference between different trunking configuration.

6. PRACTICAL CALCULATIONS

One of the objectives of traffic theory investigations is the development of practical dimensioning techniques. In this respect, the present paper has demonstrated that there is no significant difference in accuracy between the ER model and the I.P.P., so that either can be used with equal confidence and the choice is merely a matter of convenience of calculations. Also, it has been shown that there is some error involved in the use of either and that simplifying approximations may be permissible.

In calculations where the overflow route has full availability the ER model is fairly easy to use, and lends itself to a recursive, circuit by circuit, calculation using the formula.

$$U(x) = A\,(c+x)/(U(x-1)+x+c) \qquad \text{..........(1)}$$

Starting with

$$U(c) = M$$

Where:- M is the mean of the traffic offered to the common route.
A,c are the ER parameters of this traffic
U(x) is the traffic lost from a common route of x circuits.

The most time consuming process in using the ER model is in determining the ER parameters, which cannot be obtained explicitly from M & V. The A.P.O. adopts an iterative process, starting with the following approximation by Rapp (Ref.4).

$$\left.\begin{aligned} A &= V+3(V/M)(V/M-1) \\ c &= A/(1-1/(M+V/M)-M-1) \end{aligned}\right\} \text{........ (2)}$$

It now appears that sufficient accuracy can be obtained by using these approximations directly in equations (1). This is not the same as using an ER model since equation (1) assumes that M is a particular function of A & c, and this relationship does not hold for the approximations. Nevertheless, the procedure gives exact values for the traffic lost from a single circuit, and quite small errors for any number of circuits.

This is a very simple process for computation, and faster than any possible method based on the I.P.P. model. The only more economical method would be one based on curve fitting, such as that described by Berry (Ref.5). In the past such methods have tended to be regarded as less accurate than the ER model, but on the evidence here presented this may not be so.

In cases where the overflow route has internal blocking the ER model is less satisfactory and a variety of procedures is used, most of them based initially on the ER model. These are all either considerably slower than the full availability ER model, or use approximations whose effect is rather difficult to evaluate.

In these situations the I.P.P. model is comparatively easy to use, and considerably faster than the current A.P.O. method. Appendix I gives the derivation of the necessary traffic formulae for a generalised route with internal blocking. One practical difficulty is that the I.P.P. model has three parameters, and the usual representation of overflow traffic provides only the mean and variance, so that it is necessary to devise a role for setting the third parameter. Details of one possible process are given in appendix 2.

7. EFFECTS OF HIGHER MOMENTS

Attempts were made to develop models which made use of the higher moments to increase the accuracy. Although these were inconclusive they led to some interesting results.

One obvious line of investigation was to fit the I.P.P. model to the first three moments of an overflow distribution. Paradoxically this invariably gave less accurate approximation than an I.P.P. fitting the moments of the ER model. One such curve is shown in figure 4, (curve "C") and it can be seen that it diverges in the wrong direction. Note also that the traffic carried by the first circuit is unchanged. This, in fact is a particular case of a relationship which was found to apply to the moments of the traffic offered and lost from full availability routes offered Poisson, or interrupted Poisson traffic. This relationship is

$$\frac{1}{F(x,c)} = \frac{1}{F(x,c-1)} + \frac{1}{F(x+1,c-1)}$$

where F(x,c) is the xth factorial moment of the traffic overflowing from c circuits.

Setting x=1, c=1 gives $F(1,1) = 1/(1/F(1,0)+1/F(2,0))$ showing that, for these types of traffic, the overflow from one circuit is dependent only on the first two moments of the offered traffic. Clearly, it if applies to poisson and interrupted poisson traffic it also applies to the overflow of these types of traffic from a full availability route.

(The relationship can be demonstrated for poisson traffic by algebraic manipulation of the results of Riordan (Ref.6) and for interrupted poisson the factorial moments can be derived by a similar process.)

It is manifest that for the combined overflows of cases 2 to 6 the same relationship does not apply, and that therefore both the I.P.P. and the ER model differ from actual overflow traffic in a fundamental manner. Therefore, any more accurate model must be of a different type. A characteristic of the ER and the I.P.P. models is that they switch between states for which the birth co-efficients are zero, and a state for which the birth co-efficient is a constant value. In contrast to this, the combination of several overflows can offer traffic to the combined route at various intensities, depending on the sources which at that instant have all circuits busy and are therefore, offering traffic. Any more elaborate model will probably need to include this feature. Such models were investigated by Palm (Ref.7) and others, but led to extremely difficult calculations.

Cases B & C in Table 3 illustrate the range of models which can generate traffic of the same moments. They are the two negative binomial models designated 3.4.1 and 4.4.1 in Wallstroms paper (Ref. 8). The second model, in which the arrival rate is a function of the calls carried on the route to which it is offered is usually regarded as an artificial approximation. However, it is identical to an I.P.P. model in which $\ell \to \infty$, and therefore is no less artificial than the negative binomial itself.

8. CONCLUSIONS.

This investigation has -

(1) given a clearer indication of the nature of the errors in the use of the ER model.

(2) shown that an I.P.P. model fitting the first three moments of the ER model is equally accurate and more easily used in some cases.

(3) shown that greater accuracy, at far greater cost, is possible by using I.P.P. models to replace each of the components of the overflow traffic.

ACKNOWLEDGEMENT.

The author wishes to acknowledge the assistance of Mr. J. Sewell who collaborated in earlier work on the interrupted poisson process from which this paper was developed. In particular he derived algorithms for calculating the moments of the overflow, when I.P.P. traffic is offered to a route with blocking. Appendix 1 is based largely on this work.

REFERENCES

(1) WILKINSON, R.I. Theories for toll traffic engineering in the U.S.A. B.S.T.J. Vol 35, p 421, March 1956.
(2) KUCZURA, A. The interrupted poisson process as an overflow process. B.S.T.J. Vol 52 pp 437-438, March 1973.
(3) KIBBLE, W.F. The determination of telephone traffic blocking in groups of trunk circuits. G.E.C.-A.E.I. Telecommunications No.37, pp11-22, 1968.
(4) RAPP, Y. Planning of junction networks in a multi-exchange area. Ericsson Tecnics 1/1964 and 2/1965.
(5) BERRY, L.T.M. An explicit formula for dimensioning links offered overflow traffic. Australian Telecommunication Research Vol 8, No.1, pp13-17, 1974.
(6) RIORDAN, J. Derivation of moments of overflow traffic. Appendix to Ref (1).
(7) PALM, C. Intensitätsschwankungen im Fernsprechverkehr. Ericsson Tecnics, 1943:44 pp3-189.
(8) WALLSTRÖM, B. Congestion studies in telephone systems with overflow facilities. Ericsson Tecnics 3/1966.

APPENDIX 1

INTERRUPTED POISSON TRAFFIC CALCULATIONS

The interrupted poisson model has two states as shown in figure 5. In state 1, it is a source of poisson traffic of intensity l, while in state 0 no traffic is generated. The transition probabilities from state 0 to State 1 is w/rdt, and from state 1 to state 0 is g/rdt where r is the average holding time.

The traffic problem is to compute the lost traffic and grade of service when such traffic is offered to a route. Assume a generalised route of c circuits with blocking factors W(n) where W(n) is the probability that no outlet is accessible when n circuits are occupied. W(c) of course is 1.

When interrupted poisson traffic is offered to this route, there are 2C+2 states and the transition probabilities are shown, normalised in the state diagram in figure 6.

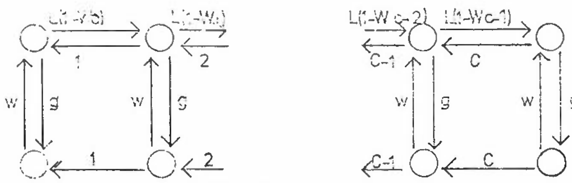


FIGURE 6

For a route of 3 circuits the coefficients of the state equation are:-

P_{00}	P_{01}	P_{10}	P_{11}	P_{20}	P_{21}	P_{30}	P_{31}	
w	-g	-1						=0
-w	g+ℓ(1-Wo)	0	-1					=0
		w+1	-g	-2				=0
	-ℓ(1-Wo)	-w	g+1+ ℓ(1-W1)	0	-2			=0
				w+2	-g	-3		=0
			-ℓ(1-W1)	-w	g+2+ ℓ(1-W2)	0	-3	=0
						w+3	-g	=0
					-ℓ(1-W2)	-w	g+3	=0

Any seven of these together with the normalising equation form a set from which the state probabilities can be obtained. The best procedure is to delete the eighth equation and simplify the set by replacing the second, fourth and sixth by the sum of the first two, the first four and the first six respectively. This gives the following equations.

P_{00}	P_{01}	P_{10}	P_{11}	P_{20}	P_{21}	P_{30}	P_{31}	
w	-g	-1						=0
	ℓ(1-W0)	-1	-1					=0
		w+1	-g	-2				=0
			ℓ(1-W1)	-2	-2			=0
				w+2	-g	-3		=0
					ℓ(1-W2)	-3	-3	=0
						w+3	-g	=0
1	1	1	1	1	1	1	1	=1

From the first seven equations it is possible to express all the state probabilities as multiples of P_{31}, or in the general case of c circuits, as multiples of $P_{c,1}$ i.e.:-

$$P_{n,m} = A_{n,m} * P_{c,1}$$

The multiples are given by:-

$$A_{c,1} = 1$$

$$A_{c,0} = g/(w+c)$$

$$A_{n,1} = (n+1)(A_{n+31,o} + A_{n+1,0})/(\ell(1-W_n))$$

$$A_{n,0} = ((n+1)(A_{n+1,0})+gA_{n,1})/(w+n)$$

Let $S = \sum_{n=0}^{c} \sum_{m=0}^{1} A_{n,m}$

then $P_{n,m} = A_{n,m}/S$

and the lost traffic is given by :-

$$U = (W_o P_{o,1} + W_1 P_{1,1} + W_2 P_{2,1} + + \; + P_{c,1})$$

and the grade of service by :-

$$B = U/\text{Offered Traffic}$$

$$= (W_o P_{o,1} + W_1 P_{1,1} + \; + \; + P_{c,1})(w+g)/w$$

The above expressions are easily programmed on a computer, and the processing time is two or three times as long as for poisson traffic offered to a similar route.

For full availability the W_n are all equal to zero except for W_c=1 and the one sub-routine can be used for both cases. It is also possible in the full availability case to develop an explicit formula which is useful for further investigations but less economical to compute.

A difficulty with the I.P.P. model is that it has three parameters and therefore requires three conditions to be specified. Two of these are that the model should match the first two moments of the actual distribution but matching the third moment also is undesriable since this makes the model less accurate than the ER model.

Ref(2) gives a method which makes the first three moments match those of the corresponding ER lmodel, but this is a rather complicated procedure. It also gives another method in which ℓ is specified by an empirical formula and the remaining parameters used to match the first two moments. A rather batter formula than that given in ref(2) is:-

$\ell = V+2(V/M)(V/M-1)$
from which
$v=g+w= (V/M-1)/(\ell-M+1-V/M)$
$w= vM/\ell$
$g=v-w$

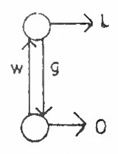


FIGURE 5 I.P.P. MODEL

APPENDIX 2

NOTES ON NUMERICAL METHODS

This appendix lists formulae and procedures which were used in the development of this paper. None are original, except for some algebraic manipulations to make them more convenient to implement on a computer.

MOMENTS OF OVERFLOW TRAFFIC

Second and higher moments are most conveniently calculated as factorial moments, and converted to other forms if desired. The formulae given by Riordan & Kosten can be shown to be equivalent to the following recursion.

$$\frac{1}{F(x+1)} = \frac{1}{x}\left(\frac{1}{AF(x-1)} - \frac{1}{F(x)} + \frac{N+x}{AF(x)}\right)$$

for the overflow traffic arising fromA erlangs offered to

N trunks and Fx is the xth factorial moment.

Note that for x = 1.

$$\frac{1}{F_2} = \frac{1}{AF_o} - \frac{1}{F_1} + \frac{N+1}{AF_1}$$

$$\text{or } F2 = \frac{AF1}{F_1 - A+N+1}$$

which is in the form given by Riordon.

FACTORIAL MOMENTS OF A COMBINED DISTRIBUTION

The factorial moments of the sum of two distributions can be calculated by converting to cumulants, adding, and converting back. However the arithmetic is extremely messy except for the first two or three moments.

The factorial moments can be used directly in the following equation.

$$Fn,a+b = Fn,a+n(Fn-1,a),(F1,b) + \frac{n(n-1)}{2}(Fn-2,a)(F2,b)+ - - - - + Fn,b$$

where Fn, a + b is the nth factorial moment of the sum of two distributions 'a' & 'b'

Fna is the nth factorial moment of distribution 'a'

Fnb is the nth factorial moment of distribution 'b'.

Discussion

A. KUCZURA, U.S.A. : *In your Figure 3, I.P.P. Accurate Model, it seems to me that you could obtain by this method the blocking seen by the individual streams. Have you done so? If you have, what results, with respect to accuracy, can you quote.*

A.H. FREEMAN, Australia : *This is correct, and indeed it is necessary to calculate the lost traffic of the individual streams and add them to obtain total losses. These have been compared with 28 of the exact results of Kibble (Ref. 3). For the 56 individual streams the ratios (G.O.S. I.P.P./GOSEXACT) had a mean value of .995 and standard deviation .012.*

E. JENSEN, Spain : *It is known that assuming a batch-poisson arrival process with geometrically distributed batch sizes to an infinite full availability group, the resulting stationary State distribution in the group will be negative binomial, provided that holding times are neg. exp. distributed.*

Now, considering that the neg. binomial distribution is a two-moment fit to the exact stationary State distribution in an infinite overflow group, if the traffic offered to the primary group is poisson, one might be led to the possibility of substituting the physical overflow arrival process by a batch-poisson process with a suitable batch size distribution. My question then is whether you have made any approach in this direction.

A.H. FREEMAN, Australia : *No attempts have been made to use the above model. An initial examination suggests that it would involve rather complex calculating procedures.*

K. WILSON, Australia : *In your paper you've considered mainly cases where means and variances are the same for each stream.*

I've noticed in my work that there are significant differences between this case and the case where the means are different and in fact the further the means depart from the equal case the greater the difference in behaviour. Have you done further work on non equal cases and what were your findings.

A.H. FREEMAN, Australia : *In Figure 4, cases 2, 4, 5 and 6 are mixtures of traffic of the same mean to variance ratio and suggest that the main factor in these cases is the number of sources, whether or not they are equal. Case 3 is the only one with a mixture of sources of different variance to mean ratio and, as you point out, the behaviour is markedly different.*

BIOGRAPHY

A.H. FREEMAN, who is a member of the Institution of Engineers, Australia, joined the A.P.O. in 1938 as a cadet draftsman in Sydney and was promoted to engineer in 1946. Until 1956 he was employed in the Radio Section, mainly on the installation and operation of broadcasting transmitters.

He then transferred to the Transmission Planning Section, at the time when the basic plans for introducing common control switching systems, and for a fully automatic trunk network were evolving, and participated extensively in the work which produced the "Community Telephone Plan 1960". During this period he was in charge of the "COMET" project which was the first application of electronic computers to Telephone network design in Australia. He was chosen as an APO representative at the fourth "International Teletraffic Congress" in London in 1964 where he presented a paper on the results of this Study.

From 1966 to 1972 he was Supervising Engineer, Trunk Service and Telegraphs, responsible for oversight of equipment maintenance in country areas of New South Wales and for installation and maintenance of Telegraph and Data facilities. Since 1972 he has been a Supervising Engineer in the Planning and Programming Branch, N.S.W.

Correlation Induced in Traffic Overflowing from a Common Link

K. G. WILSON

South Australia Institute of Technology, Adelaide, Australia

ABSTRACT

Certain links in alternative routing networks are offered independent streams of traffic from two or more sources. Previously, the behaviour of this traffic has been predicted using simulation or the equivalent random method. An analytic model for this case has been developed and can be used as a tool in investigating traffic behaviour in alternative routing networks in general.

This paper considers the moments of the traffic overflowing from the shared link and in particular the covariance. The covariance is often assumed to be zero, e.g. in simulation or else bypassed as in the equivalent random method which considers the offered traffic as if it were only one stream.

An iterative solution to the problem is given and an analytic solution to a special case, in which the offered traffic is random, is derived.

1. INTRODUCTION

In early practical models of telephone networks the traffic was described by its mean alone. More recently both mean and variance have been considered. A commonly made assumption has been that streams offered to a common link are independent, and hence the variance of the offered traffic is simply the sum of the variances of the separate streams. However, whenever the offered streams are overflows from routes sharing a common link the assumption is no longer valid.

In order to investigate the correlation between such streams the following simple network was considered (fig.1).

The network has one origin exchange "0", two destination exchanges "D1" and "D2" and two tandem exchanges X_1, Y_1.

The routes the traffic may take are described below.

O-D pair	1st	2nd	3rd
1 (0 & D1)	1	3,5	4,5
2 (0 & D2)	2	3,6	4,6

Table 1. Alternative Routes

Two independent streams of traffic originated at "0"; one consisting of calls intended for D1, the other for D2. These streams have Poisson distributions with means a_1 and a_2 respectively.

The 1st choice (direct) routes have d_1, d_2 junctions respectively and link 3, the common link on the second choice routes has c junctions.

It is assumed that links 5 and 6 have enough junctions to carry all traffic offered to them and hence any congestion will occur on links 3 and 4. It is assumed that there are enough junctions between the tandem exchanges X_1 and Y_1 to carry any calls offered to that link.

The traffic overflowing from 3, in particular the means, variances and covariance of this traffic, is of interest and from these moments the traffic actually carried on link 3 can also be obtained.

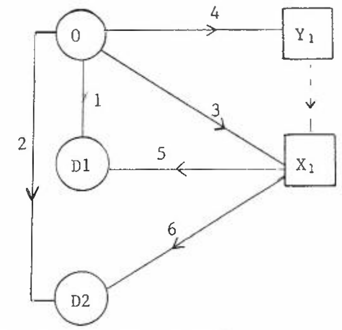


Figure 1. The Network

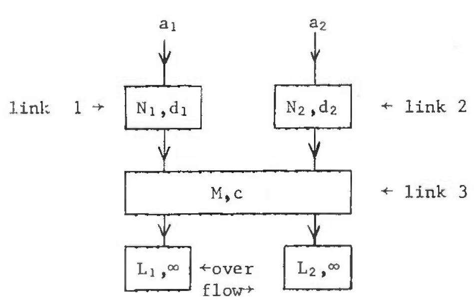


Figure 2. System Configuration

Where $\boxed{x,y}$ – x = number of busy servers
y = total number of servers.

2. MATHEMATICAL MODEL

The development of this model is an adaption of the approach used by Neal [1].

The network can be considered as a system of servers, each junction in a link being one server. Since the traffic overflowing from 3 is important, the last "blocks" of servers are infinite and there is no loss from the system.

Calls arrive at the direct link servers independently with Poisson distributions, means a_1, a_2 respectively. The mean of the total number of arrivals is $a = a_1 + a_2$. Service times throughout the system have independent negative exponential distributions with unit mean. Service times correspond to the holding time of a call.

It is assumed that the system has reached equilibrium. Then the state probability density function is

$$f(n_1,n_2,m,\ell_1,\ell_2) = P(N_1=n_1,N_2=n_2,M=m,L_1=\ell_1,L_2=\ell_2) \quad (1)$$

and f=0 outside the range $0 \le n_i \le d_i$, $0 \le m \le c$, $\ell_i \ge 0$.

The state equations for f have the following form.

For $0 \le n_i \le d_i - 1$, $0 \le m \le c$, $\ell_i \ge 0$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)f(n_1,n_2,m,\ell_1,\ell_2) &= a_1 f(n_1-1,n_2,m,\ell_1,\ell_2) \\
&+ a_2 f(n_1,n_2-1,m,\ell_1,\ell_2) \\
&+ (n_1+1)f(n_1+1,n_2,m,\ell_1,\ell_2) \\
&+ (n_2+1)f(n_1,n_2+1,m,\ell_1,\ell_2) \\
&+ (m+1)f(n_1,n_2,m+1,\ell_1,\ell_2)^* \\
&+ (\ell_1+1)f(n_1,n_2,m,\ell_1+1,\ell_2) \\
&+ (\ell_2+1)f(n_1,n_2,m,\ell_1,\ell_2+1)
\end{aligned} \tag{2}
$$

*this term is zero when m=c

There are 6 other equations corresponding to the boundary conditions.

This infinite set of equations can be simplified by introducing a two dimensional binomial moment generating function B.

$$
B(n_1,n_2,m;x_1,x_2) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} f(n_1,n_2,m,k_1,k_2)(1+x_1)^{k_1}(1+x_2)^{k_2} \tag{3}
$$

for $0 \le n_i \le d_i$, $0 \le m \le c$, $-1 \le x_i \le 0$

$$
= \sum_{\ell_1=0}^{\infty}\sum_{\ell_2=0}^{\infty} B_{\ell_1,\ell_2}(n_1,n_2,m)x_1^{\ell_1}x_2^{\ell_2} \tag{4}
$$

where $B_{\ell_1,\ell_2}(n_1,n_2,m) = \sum_{k_1=\ell_1}^{\infty}\sum_{k_2=\ell_2}^{\infty} \binom{k_1}{\ell_1}\binom{k_2}{\ell_2} f(n_1,n_2,m,k_1,k_2)$ (5)

are the binomial moments.

Define

$$
B_{(\ell_1,\ell_2)} = \sum_{n_1=0}^{d_1}\sum_{n_2=0}^{d_2}\sum_{m=0}^{c} B_{\ell_1,\ell_2}(n_1,n_2,m) \tag{6}
$$

$$
= E\left[\binom{L_1}{\ell_1}\binom{L_2}{\ell_2}\right] \tag{7}
$$

In particular

$$
\begin{aligned}
B_{(1,0)} &= E(L_1) && (8)\\
B_{(0,1)} &= E(L_2) && (9)\\
B_{(1,1)} &= E(L_1 L_2) && (10)\\
B_{(2,0)} &= \tfrac{1}{2}E(L_1^2-L_1) && (11)\\
B_{(0,2)} &= \tfrac{1}{2}E(L_2^2-L_2) && (12)
\end{aligned}
$$

Hence means, variances and covariances of L_1, L_2 can be calculated from these moments.

A system of equations in the binomial moments can be derived from the state equations by multiplying by $(1+x_1)^{\ell_1}(1+x_2)^{\ell_2}$, summing over ℓ_1,ℓ_2 and equating like powers of x_1 and x_2.

The following abbreviations are used

$$
\begin{aligned}
B(,,,) &= B_{\ell_1,\ell_2}(n_1,n_2,m)\\
B(x,,) &= B_{\ell_1,\ell_2}(x,n_2,m)\\
B(,y,) &= B_{\ell_1,\ell_2}(n_1,y,m) \quad \text{etc.}\\
B_{x,}(,,,) &= B_{x,\ell_2}(n_1,n_2,m) \quad \text{etc.}
\end{aligned}
$$

For $0 \le n_i \le d_i - 1$, $0 \le m \le c$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ (n_1+1)B(n_1+1,,)+(n_2+1)B(,n_2+1,)\\
&+ (m+1)B(,,m+1)
\end{aligned} \tag{13}
$$

For $0 \le n_2 \le d_2 - 1$, $0 \le m \le c-1$, $n_1 = d_1$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,)+a_2B(,n_2-1,)\\
&+ a_1B(,,m-1)+(n_2+1)B(,n_2+1,)\\
&+ (m+1)B(,,m+1)
\end{aligned} \tag{14}
$$

For $0 \le n_1 \le d_1 - 1$, $0 \le m \le c-1$, $n_2 = d_2$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ (n_1+1)B(n_1+1,,) + a_2B(,,m-1)\\
&\quad (m+1)B(,,m+1)
\end{aligned} \tag{15}
$$

For $0 \le n_2 \le d_2 - 1$, $n_1 = d_1$, $m=c$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ a_1B(,,m-1) + (n_2+1)B(,n_2+1,)\\
&+ a_1B(,,,) + a_1B_{\ell_1-1,}(,,,)
\end{aligned} \tag{16}
$$

For $0 \le n_1 \le d_1 - 1$, $n_2 = d_2$, $m=c$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ (n_1+1)B(n_1+1,,) + a_2B(,,m-1)\\
&+ a_2B(,,,) + a_2B_{,\ell_2-1}(,,,)
\end{aligned} \tag{17}
$$

For $0 \le m \le c-1$, $n_i = d_i$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ a\,B(,,m-1)\\
&+ (m+1)B(,,m+1)
\end{aligned} \tag{18}
$$

For $n_i = d_i$, $m=c$

$$
\begin{aligned}
(a+n_1+n_2+m+\ell_1+\ell_2)B(,,,) &= a_1B(n_1-1,,) + a_2B(,n_2-1,)\\
&+ a\,B(,,m-1) + a\,B(,,,)\\
&+ a_1B_{\ell_1-1,}(,,,) + a_2B_{,\ell_2-1}(,,,)
\end{aligned} \tag{19}
$$

Let $B_{\ell_1,\ell_2}(n_1,n_2,m) = 0$ outside range $0 \le n_i \le d_i$, $0 \le m \le c$ $1_i \ge 0$.

Summing (13) - (19) over the respective ranges gives

$$
(\ell_1+\ell_2)B_{(\ell_1,\ell_2)} = a_1\sum_{n_2=0}^{d_2} B_{\ell_1-1,\ell_2}(d_2,n_2,c) + a_2\sum_{n_1=0}^{d_1} B_{\ell_1,\ell_2-1}(n_1,d_2,c) \tag{20}
$$

Using (20) all the moments in (8) - (12) can be calculated if $B_{\ell_1,\ell_2}(n_1,n_2,m)$ is known for all (n_1,n_2,m) and for $(\ell_1,\ell_2) = (0,0)$, $(0,1)$ and $(1,0)$.

Equations (13) - (19) can be written as a single matrix equation

$$
M\underset{\sim}{b} = \underset{\sim}{g} \tag{21}
$$

where $B_{\ell_1,\ell_2}(n_1,n_2,m)$ will be the r^{th} element of $\underset{\sim}{b}$ for $r = (n_1+1) + n_2(d_1+1) + m(d_1+1)(d_2+1)$.

M is square, size $(d_1+1)(d_2+1)(c+1)$.

(21) can be rewritten

$$
(I-A)\underset{\sim}{b} = \underset{\sim}{f} \tag{22}
$$

or

$$
\underset{\sim}{b} = A\underset{\sim}{b} + \underset{\sim}{f}. \tag{23}
$$

Equation (23) can be solved using Gauss-Siedel iteration with successive over relaxation.

The matrix A is highly structured and in fact has at most 7 non zero elements per row, in 7 diagonal bands. A program was written to solve this equation and hence calculate the overflow moments.

Note that when $\ell_1=\ell_2=0$ the matrix M is singular, ((20) reduces to 0=0). Since $B_{0,0}(n_1,n_2,m) = P(N_1=n_1, N_2=n_2, M=m)$, the last row of M is replaced by a row of 'ones' and the last element of $\underset{\sim}{g}$ by a 'one', to give the matrix M full rank.

3. A SPECIAL CASE - NO DIRECT JUNCTIONS

When there are no direct junctions $n_i=d_i=0$ and (13) - (19) reduce to -

For $0 \le m \le c-1$, $\ell_i \ge 0$

$$
(a+m+\ell_1+\ell_2)B_{\ell_1,\ell_2}(m) = a\,B_{\ell_1,\ell_2}(m-1) + (m+1)B_{\ell_1,\ell_2}(m+1) \tag{24}
$$

$$
(c+\ell_1+\ell_2)B_{\ell_1,\ell_2}(c) = a\,B_{\ell_1,\ell_2}(c-1) + a_1B_{\ell_1-1,\ell_2}(c) + a_2B_{\ell_1,\ell_2-1}(c) \tag{25}
$$

define $\beta=\beta(t)=\sum_{m=0}^{\infty}B_{\ell_1,\ell_2}(m)t^m$. (26)

(24), (25) →

$$(a(1-t)+\ell)\beta=(1-t)\frac{d\beta}{dt} \tag{27}$$

where $\ell=\ell_1+\ell_2$

$$\beta(t) = K\frac{e^{at}}{(1-t)^{\ell}} \tag{28}$$

Following Riordan [2], define $\sigma_\ell(m)$ by

$$\sum_{m=0}^{\infty}\sigma_\ell(m)t^m = \frac{e^{at}}{(1-t)^{\ell}} \tag{29}$$

(26), (28) and (29) →

$$B_{\ell_1,\ell_2}(m) = K\,\sigma_\ell(m) \tag{30}$$

and $K = \beta(0) = B_{\ell_1,\ell_2}(0)$ (31)

Hence (25) can be written

$$(c+\ell)B_{\ell_1,\ell_2}(0)\sigma_\ell(c) = a\,B_{\ell_1,\ell_2}(0)\sigma_\ell(c-1)$$
$$+\ \sigma_{\ell-1}(c)(a_1B_{\ell_1-1,\ell_2}(0)+a_2B_{\ell_1,\ell_2-1}(0)) \tag{32}$$

but $(c+\ell)\sigma_\ell(c) - a\sigma_\ell(c-1) = \ell\sigma_{\ell+1}$ (33)

from Riordan [2], (p.510), and hence

$$B_{\ell_1,\ell_2}(0) = \frac{1}{\ell}(a_1B_{\ell_1-1,\ell_2}(0)+a_2B_{\ell_1,\ell_2-1}(0))\frac{\sigma_{\ell-1}(c)}{\sigma_{\ell_2+1}(c)} \tag{34}$$

now $\sum_{m=0}^{c}B_{0,0}(m) = 1$ (35)

$$\rightarrow 1 = \sum_{m=0}^{c}B_{0,0}(0)\sigma_0(m) = B_{0,0}(0)\sigma_1(c) \tag{36}$$

c.f. Riordan [2]

$$\rightarrow B_{0,0}(0) = 1/\sigma_1(c) \tag{37}$$

since $B_{0,-1}(m) = B_{-1,0}(m) = 0$

and $\frac{\sigma_0(c)}{\sigma_1(c)} = E_c(a)$ (the Erlang loss formula) (38)

$$\frac{\sigma_1(c)}{\sigma_2(c)} = \frac{1}{c+a+aE_c(a)+1} = \frac{1}{D}\ \text{(say)} \tag{39}$$

(34), (30) and (31) give

$$B_{1,0}(m) = a_1B_{0,0}(0)\,\frac{\sigma_0(c)}{\sigma_2(c)}\cdot\sigma_1(m) \tag{40}$$

hence $B_{1,0}(c) = a_1\,\frac{\sigma_0(c)}{\sigma_1(c)}\,\frac{\sigma_1(c)}{\sigma_2(c)}$ (41)

$$= \frac{a_1E_c(a)}{D} \tag{42}$$

and $B_{0,1}(c) = \frac{a_2E_c(a)}{D}$ (43)

Now (8) - (12) can be evaluated using (20)

$$B_{(1,0)} = a_1B_{0,0}(c) = a_1E_c(a) \tag{44}$$
$$B_{(0,1)} = a_2B_{0,0}(c) = a_2E_c(a) \tag{45}$$
$$B_{(1,1)} = \tfrac{1}{2}(a_1B_{0,1}(c) + a_2B_{1,0}(c))$$
$$= \frac{a_1a_2E_c(a)}{D} \tag{46}$$
$$2B_{(2,0)} = a_1B_{1,0}(c) = a_1^2\frac{E_c(a)}{D} \tag{47}$$
$$2B_{(0,2)} = a_2B_{0,1}(c) = a_2^2\frac{E_c(a)}{D} \tag{48}$$

and hence the overflow moments are

$$m_1 = a_1E_c(a) \tag{49}$$
$$m_2 = a_2E_c(a) \tag{50}$$
$$m_t = aE_c(a)\quad\text{(total mean)} \tag{51}$$
$$v_1 = 2B_{(2,0)} + B_{(1,0)} - B^2_{(1,0)}$$
$$= m_1(1-m_1 + \frac{a}{c-a+m_t+1}) \tag{52}$$

$$v_2 = m_2(1-m_2 + \frac{a_2}{c-a+m_t+1}) \tag{53}$$

$$\text{cov} = B_{(1,1)} - B_{(1,0)}B_{(0,1)}$$
$$= m_1(-m_2 + \frac{a_2}{c-a+m_t+1}) \tag{54}$$

since $m_1a_2 = m_2a_1$ the following expressions are equivalent

$$\text{cov} = m_1(-m_2 + \frac{a_2}{c-a+m_t+1})$$
$$= m_2(-m_1 + \frac{a_1}{c-a+m_t+1}) \tag{55}$$
$$= \tfrac{1}{2}m_1(-m_2 + \frac{a_2}{c-a+m_t+1})$$
$$+ \tfrac{1}{2}m_2(-m_1 + \frac{a_1}{c-a+m_t+1}) \tag{56}$$

(56) being a symmetric expression for the covariance.

This analysis confirms the assumption that when 2 random streams are offered to a single link, the overflow mean is proportional to the offered mean. Further, it gives exact formulas for variance and covariance.

The formulas generalise to the case of r independent random streams offered to a single link, as follows,

for $a = \sum_{i=1}^{r} a_i$, $m_t = \sum_{i=1}^{r} m_i$

$$m_i = a_iE_c(a) \tag{57}$$
$$v_i = m_i(1-m_i + \frac{a_i}{c-a+m_t+1}) \tag{58}$$
$$\text{cov}_{ij} = \text{cov}_{ji} = m_i(-m_j + \frac{a_j}{c-a+m_t+1}) \tag{59}$$

Finally, the structure of the formulas for m_i and v_i are very similar to those for the total mean and variance (corresponds to the case where a single random stream, mean a, is offered to a single link).

$$m_t = aE_c(a) \tag{60}$$
$$v_t = m_t(1-m_t + \frac{a}{c-a+m_t+1}) \tag{61}$$

4. RESULTS

The model has 5 parameters, viz. a_1, a_2, d_1, d_2 and c and an investigation of the effects of varying all of these, requires a large amount of data. This data will then be used to find an approximate formula for the overflow moments. The special case gives exact limits with which the approximations must conform.

From the data collected at this stage, the following trends have been noted, with consideration given to v_t, the variance of the combined overflow streams, and R, the relative error in assuming independence.

viz. $R = \frac{v_t-(v_1+v_2)}{v_t}$

$$= \frac{2\text{cov}}{v_t}$$

(R is a measure of the proportion of the total variance contributed by the covariance.)

If the number of junctions on the common link is increased, and all else is fixed, then v_t decreases but R increases.

If the variance of one stream offered to the common link increases, its mean and all other parameters being fixed, then v_t increases and R remains fairly constant.

If both the mean and variance of one stream increase, their ratio and other parameters being constant, then v_t increases but R decreases.

These cases seem to confirm the intuitive suspicion that the correlation in the overflow is related to the amount that the two streams are mixed, that is, the greater the ratio of carried traffic to offered traffic on the common link the greater the contribution made by the covariance to the total variance. In general, once the carried to offered ratio was greater than 1/3, R was more than 10%, being as much as 28% in some cases.

Attempts to solve the more general case analytically have so far been unsuccessful. However, it may be noted that Neal [1] did not obtain an analytic solution to his model, although he reduced the size of his problem by analytic methods. The solution of the reduced problem was still obtained numerically.

5. CONCLUSION

The model presented here is a tool to be used in understanding the behaviour of traffic streams sharing a common link in an alternative routing network. It can be used to generate a large amount of data, to enable empirical formulas for the overflow moments to be calculated and indeed may yet be solved analytically, as it has been for the special case.

Results obtained from numerical solution of the problem indicate that correlation is significant and hence the covariance must be taken into consideration in this type of network.

6. ACKNOWLEDGEMENTS

The author expresses his appreciation of the guidance and encouragement of Dr. L.T.M. Berry of the Department of Applied Mathematics, University of Adelaide. Thanks are also due to other members of the Department, particularly to Drs. C. Pearce and W. Henderson and to Dr. R. Harris and Mr. J. Rubas of the Australian Telecommunications Commission. Work for this project was supported in part by the Australian Telecommunications Commission.

7. REFERENCES

1. S. NEAL "Combining Correlated Streams of Non-Random Traffic" The Bell System Technical Journal, Vol.50, No.6 (pp2015-2037), July-August, 1971.

2. R.I. WILKINSON (Appendix by J. RIORDAN) "Theories for Toll Traffic Engineering in the U.S.A." Bell System Technical Journal, Vol.35, No.2 (pp421-514) March, 1956.

Discussion

J.P. FARR, Australia : *In Section 1 you state that tandem exchanges X1 and Y1 are usually situated in the same building and it is assumed that there are an "infinite" number of junctions between them. There is no reason why X1 and Y1 have to be in the same building and I am aware of networks in which this is the case. The number of junctions between X1 and Y1 may be dimensioned for a fixed grade of service thus reducing the amount of traffic carried. To what extent does this require your results to be modified.*

K.G. WILSON, Australia :

(i) The model may be considered as a link system in which traffic is carried on a link if there is a free junction irrespective of its ability to be carried on succeeding link.

(ii) Further I really consider traffic offered to this 3rd choice route and not the traffic carried on it.

A. KUCZURA, U.S.A. : *In solving your special case, in which the offered traffic is random, you observed that the overflow mean is proportional to the offered mean. I wonder if you had examined this relationship for your general case.*

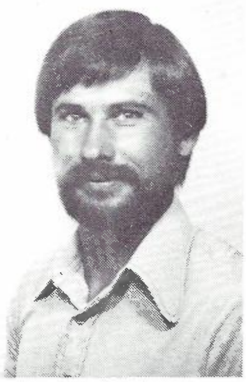
K.G. WILSON, Australia : *Not directly. I am at present attempting to obtain an approximate formula for the proportion of traffic carried from each stream. This will allow the overflow means to be calculated.*

R. SCHEHRER, Germany : *In your paper it is stated that the covariance must be taken into consideration" in a certain type of network. Now the question : Does any method exist which takes the covariance into account or what sort of method would the author suggest for this purpose.*

K.G. WILSON, Australia : *Looking at the properties of separate streams offered to a common link seems one more step in the understanding of teletraffic. When this approach is used, covariance is one of the moments of interest the others being the mean and variance of the separate overflow streams and the mean traffic from each stream which is actually carried. I know of no models which use covariance, probably because there is no simple method of calculating it at present.*

Dr. Harris, Aust. Telecom is looking at a model in which he uses the variances of the separate overflow streams.

Dr. Olsson obtained an approximation for the ratio of the means of the overflow traffic. Prof. Wallström is also looking at this problem. Dr. Berry, Aust. wishes to know ratio of carried traffic to help understand chain flows on a Network.



BIOGRAPHY

KYM WILSON completed an Honours degree in Statistics at Adelaide University in 1973. In the three years since then he has been working for a Ph.D in the Department of Applied Mathematics. For the first two years the research was partially supported by a contract with the Australian Post Office. During the third year he was a mathematics tutor at the S.A.I.T. He submitted his thesis in February 1977 and is now a petroleum engineer with Shell.

Traffic Studies of Message Switching Computers

L. D. THUAN
Control Data Australia, Melbourne, Australia
R. E. BOGNER
University of Adelaide, Adelaide, Australia

ABSTRACT

This paper studies prediction of the performance of computer controlled store and forward message switching systems, and hence the traffic levels and CPU utilisations which the systems can handle. The study is carried by both analytical and simulation methods, and the analytical results are found to be less accurate. The simulation is also used to study the dependence of response time on throughput.

1. INTRODUCTION

A computer controlled store and forward message switching system receives messages from its input terminals and stores them until they can be transmitted to the output terminals. Messages are received character by character and until a special character signifying the end of a message is received the transmission of the message to its destination terminal is delayed. The transmission is further delayed if the recipient terminal is unable to receive the message for any reason such as being busy in receiving another message.

The delay between the receipt of the end of message character and the transmission of the first character of the message to the destination terminal is termed Cross-Office Delay (COD). One of the prime objectives of system design and dimensioning is that the COD's of 90 or 95% of all messages fall within some specified upper limit. This requirement imposes restrictions on the traffic (i.e. messages per second handling capacity of the system.

The capacity of a computer controlled system may be simply described as the maximum percentage utilisation of the central processing unit (CPU) at which the performance of the system is acceptable. The performance is characterised by parameters such as COD's, number of polls (i.e. invitations to input message) per terminal per second, main store buffers utilisation, etc.

In order to plan ahead and forecast traffic growth capabilities it is necessary to estimate the capacity of the system.

Several techniques have been developed to estimate the capacity of a computer controlled system. These can be broadly categorised into three classes, namely analytical, simulation and measurement methods. In the last, the CPU utilisation, COD's and other parameters are measured on the actual system at various traffic levels. The traffic is either actually generated from the suitable number of terminals or simulated through some software means. A working system (i.e. fully implemented) is premature for these techniques which cannot be used during the system design phase.

On the other hand, analytical methods based largely on queuing theory can be used during the design of the system. However, because of assumptions needed to produce models amenable to analysis, the accuracy of the results falls short of that of the measurement methods. In a real system both methods are used depending upon various stages in the system implementation.

Commonly used analytical methods to predict performance of computer controlled stored and forward message switching systems such as Australian Post Office's Common User Data Network (CUDN) are discussed. Data relevant to this study were observed and provided by Telecom Australia; these are presented in Table 1 and Table 2. Accuracy of the results is compared with those of simulation.

The model used in both the analytical and simulation methods represents the essentials of a computer controlled store and forward system as described below.

2. MODEL OF SWITCHING SYSTEMS

To study the behaviour of a switching centre (node), and hence of the network, the node may be modelled as a black box which has several incoming lines over which messages are received. These messages are either being served or waiting to be served. After completing the servicing of these messages, the node sends them on the outgoing lines.

The model to describe the way in which switching systems resolve conflicting requests for attention of the CPU is shown in Figure 1.

This model consists of a single resource (CPU) and a system of queues which holds those messages awaiting service. In addition, there exists a scheduling algorithm, which is a set of decision rules determining which message will next be serviced. Thus a newly entering request is placed in the system of queues, and, when the scheduling algorithm permits, is given a turn in the processing facility.

The scheduling algorithm is presented in Figure 2.

The service facility (CPU) contains N physical programs. Depending on the type of a message, several programs may be required to process it. After the execution of one program is completed, the message is passed to the next program and so on until the message is fully processed and ready to be sent out.

We allocate to the programs required for the servicing of high priority messages appropriate priority numbers, so that the priority need be associated only with the program numbers. The switching centre may now be represented by Table 3.

In Table 3, programs process a message in the order of program numbers. The priority of execution of a program is always in the order of the program numbers. A physical program is defined as the one which is required to perform a specific function e.g. one program is responsible for message routings. If the same physical program is required to serve different messages of several priorities it is assigned corresponding program numbers. Hence several program numbers may correspond to one physical program and the service facility (CPU) consists of numbered programs (P_1, P_2, ..., P_m) where the subscripts indicate the priority of these numbered programs, the smaller the subscript the higher the priority of the numbered program.

Since the CPU can execute only one program at a time these programs compete amongst each other to take control of the CPU. The model now may be considered as a single server (CPU), which has several input queues which in turn generate requests for the server. The messages continue to be fed back to the CPU until their processing is completed, i.e. the execution of an end program in a series is completed. A message is then to be sent out.

If a message requiring processing from a higher priority program arrives during the execution of a program then the lower priority program currently being executed by the CPU is suspended and the CPU is allocated to the higher priority program to commence its processing. On the completion of this processing the CPU is returned to the suspended program whose processing is resumed from the point it was suspended. This method of operation is known as Pre-emptiveResume Priority discipline.

3. ASSUMPTIONS

In order to make the model as close as possible to the real world while permitting queuing theory to be used to analyse the traffic capacity, the following assumptions are made:

(1) The service discipline is assumed to be pre-emptive resume priority in which a program of higher priority has immediate precedence over one of lower priority, including the interruption of the program in service. The interrupted program regains the service where it left off only after no higher priority programs remain in the system.

(2) The term "service distribution" will be used to denote the distribution of time occupied by a program in servicing a message. In order to permit a compromise between the two convenient extreme service distributions, exponential and constant distribution, we assume the processing time of each program follows an Erlang-m distribution [1].

Measured values of average processing times were provided by Telecom Australia [2], and are shown in Table 1.

(3) The terminals and adjacent switching centre generate incoming messages such that the time intervals between the two consecutive messages of the same priority are exponentially distributed with parameters λ_k, k = 1,2...10.

With the numbers of terminals in practical situations, the assumption that the overall message generation is a poisson process for each message class is often justified.

3.4 Figure 3 shows the system to be investigated analytically. The incoming messages which demand a service are classified into n parallel queues according to their priorities λ_i. All queues are assumed to be unlimited i.e. every incoming message will be stored and processed. This assumption is almost always fulfilled, especially in the systems with dynamic core allocation. All messages are served according to pre-emptive priority discipline and first in first out is assumed within each priority class.

3.5 The time to handle an interrupt is neglected. This assumption is reasonable because large system computers have multiple register sets and hardware for interrupt handling.

4. MATHEMATICAL ANALYSIS

Based on the above assumptions, the queuing time (time spent in the system waiting and being processed) is obtained for programs of the jth priority class from [3]:

$$T_j = \frac{1}{1-U_{j-1}} \left[b_{1j} + \frac{\sum_{i=1}^{j} \lambda_i b_{2i}}{2(1-U_j)} \right]; \qquad (1)$$

where:

$$U_j = \lambda_1 b_{11} + \lambda_2 b_{12} + \ldots + \lambda_j b_{1j}, \qquad (2)$$

with $u_0 = 0$;

λ_i = arrival rate for programs of the ith class; b_{1j}, b_{2j} are the mean and second moment of service time in the j-th priority class. The moments of the Erlang distribution are given in general [2] as:

$$b_{nj} \text{ (Erlang - m service)} = \frac{(n+m-1)!}{(m-1)!} \left(\frac{T_{js}}{m}\right)^n ;$$

T_{js} is the average service time of the j-th priority class.

The cross office delay (COD) of a message of class A is the total time taken to process all the associated programs, i.e.:

$$T_A = \sum_{j \in A} T_{Aj} \ . \qquad (3)$$

Where T_{Aj} is queuing time of j-th priority program which belongs to the above mentioned set, as found from (1).

With the Erlang - 10 (m = 10) service time distribution whose average processing times are in Table 1 and the message arrival rates as in Table 2, analytical results obtained from (3) are as shown in Figures 6, 7, 8, 9 and 10.

5. VALIDATION CHECKS

To be valid a model should satisfy the following requirements:

(1) Events should occur in the correct sequence. The simulation program provides checking of this sequence, e.g. for each message a print out is available of the time at which each event (i.e. each service commencement) occurs.

(2) The statistics corresponding to the queues of traffic and servicing collected during the simulation must reflect the known input statistics i.e. the average inter-arrival times of messages must match with the inter-arrival time generated during a simulation. The same condition must apply for service times etc.

(3) The simulation results should match the known measurable results such as CPU utilisation, and cross office delay time.

6. RESULTS AND INTERPRETATION

A misinterpretation of collected data and consequently a misunderstanding of the system could arise through:

6.1 INITIAL CONDITIONS

To eliminate the transient effect of initial conditions, one selects a deletion time and deletes data obtained prior to this time. A suitable deletion time is found by making several pilot runs and collecting data at relatively short intervals. Then one plots these sample records against time to determine a reasonable deletion time [4].

6.2 STATISTICAL FLUCTUATIONS

In all simulations containing random phenomena the question arises of how long to run an experiment so that representative averages are obtained. The answer depends on how statistically accurate we wish our results to be. The technique by G.S. Fishman [5] was used to determine the condition to end a simulation run.

Fishman's technique uses a recursive estimator of the variance of the required value, and takes into account correlations between observations. In use, the variance estimate is updated at intervals of M observations of the simulation, and the simulation is terminated when the variance, or confidence interval, is reduced to a predetermined satisfactory level. The confidence intervals are based on the assumption that the observations follow a normal distribution. The 95% confidence limits shown in Figures 5 through 9 were estimated using Fishman's method.

The simulation and analytical results of cross office delay (COD) of several messages of several classes are shown in Figures 5, 6, 7, 8 and 9. These figures show clearly that under low traffic conditions, the COD's obtained by the two methods are nearly the same. However, when the traffic input increases the results differ significantly.

Sufficient explanations are quite easy to find. In the mathematical analysis, we made the assumption that input messages which require several programs arrive at the same rate for each of the programs. This assumption fails when there is a heavy input traffic, since a CPU may be busy serving those programs with higher priority. This in turn creates a longer queuing time and hence arrival rates of all the consecutive lower priority programs are no longer independent. In Figure 5 the resultant difference between the two methods is not marked until the input ratio equals 2.0 because message A is a highest priority message. For lower priority messages as in Figures 9 and 10 the simulation COD is larger than the analytical COD even at low input traffic, and this is consistent with the explanation given above.

The queuing times of messages for each individual program were also obtained. It was found that if the pre-empted program execution time is eight to ten times greater than the time to execute the pre-empting message, the effect on the delay time is minimal.

However, if the situation is reversed i.e. pre-empting is ten times greater than pre-empted, the total delay time is increased nearly 100 percent at 60 percent CPU utilization.

Figure 10 shows the variation of COD of link A messages with CPU utilization (several of these curves were also obtained for different messages). At 70 percent of CPU utilization the COD increases drastically. At 80 percent CPU utilization, the COD becomes large and very large queues were found to be building up for lower priority messages. Thus the switching centre is overloading at a traffic volume which corresponds to 2.2 times the 1972 level.

For a given traffic level, one can obtain the cumulative probability distribution function for a message from the simulation. Hence one can determine whether the objective of system design and dimensioning is met or not i.e. 90 or 95 percent of messages must have COD's within some specified limit. Figure 11 shows the cumulative distribution function of link I/P message A.

7. RESPONSE TIME

We can define the response time as the elapsed time from the last character of message transmitted from a terminal until the last character of response message appears in the same terminal.

For example, let us study the response time of the highest priority message A with the configuration of terminals, switching centre and customer computer unit (CCU) as shown as in Figure 12.

We use the previously obtained results for the cross office delay of message A and various numbers of terminals per multiplexer (MUX), the delay times of CCU, the message length etc. to obtain the response time of top priority message A under various conditions.

The system simulated provided service via polling, and the procedures were as follows:

Under normal operating conditions, several terminals may be prepared to transmit messages at the same time. Only one can do so, and the others must wait for their turn.

The "POLLING" program is executed at N times the polling rate, where N is the number of multiplexors per line. The multiplexors are polled in sequence i.e. 1, 2, ..., N. If there is a message being transmitted from CPU to any one of the multiplexors and it is time to poll MUX K, then that poll is lost. If MUX 2 is inputting message and the CPU sends a poll to MUX 1, the reply from MUX 1 must wait.

The sequence of operation is described as follows:

(a) (i) CPU polls MUX 1.

(ii) MUX 1 looks if there is a message ready at any of the terminals connected to it.

(iii) If there is no message ready for input to the CPU, then MUX 1 sends a traffic response to the CPU. Otherwise the MUX sends the message to the CPU.

(b) To send a message to the terminal, the CPU sends a select signal to the MUX defining the terminal. If the terminal is busy (e.g. it is sending a message to the CPU), the CPU waits for a predetermined interval before sending vs the "select" signal again.

When the MUX replies to a select signal, the CPU sends a message to the terminal.

The values used in the simulations were:

message input rate: 65 messages/terminal/hour
polling rate: 12 polls/sec/MUX
average message length: 80 characters (distributed as Erlang-10)
CCU delay times: 400, 600 and 1000 ms
number of terminals per MUX: 5, 10, 15 and 20.

Figure 13 shows the response time vs throughput for a system with three multiplexors. It was found that the delay time in customer computer unit (CCU) contributes significantly to the overall response time. In Figure 13, the three different curves correspond to assumed CCU delay times of 1000 ms, 600 ms, and 400 ms. For the case of 1000 ms delay time, there is a big queue building up at the CCU and the system fails when there are 15 terminals per MUX.

Figure 14 shows the response time vs message length for the case of 600 ms CCU delay. For 10 and 15 terminals per MUX 15, the response time is increased mainly due to the propagation delay. But for 20 terminals, the response time is increased significantly due to the longer waiting time at the CCU.

8. CONCLUSION

In this paper, a queuing model was established for the store and forward message switching systems. The model was analysed by an analytical and a simulation method under the following assumptions:-

(a) Poisson arrivals.

(b) Pre-emptive - resume priority service discipline.

(c) An Erlang - 10 service distribution.

Under the low input traffic rate, the analytical results are approximately equal to the simulation results, and hence the analytical method is quite useful in the sense that it gives a first sight knowledge of the system performance. However the results of the COD's obtained by the analytical method depart from those obtained by simulation at the high input traffic rates.

The simulation results are able to determine the CPU utilisation at which the system is over loaded, and hence the input traffic level which the system can handle.

The COD results obtained by the simulation are also used to study the response time against the throughput for the hypothetical model. This study is used to determine the characteristics of the system, and hence the number of terminals per multiplexor permissible for a desired response time.

9. REFERENCES

[1] U. HERZOG "Optimal Scheduling Strategies of Real Time Computers", I.B.M. Res. Develop., Sept. 1975.

[2] TELECOM AUSTRALIA "APO's CUDN 1972 Statistical Tables".

[3] I.B.M. REPORT (F20-0007-1) Analysis of Some Queuing Models in Real Time Systems, 1969.

[4] G.S. FISHMAN "Concepts and Methods in Discrete Event Digital Simulation", John Wiley and Sons, Dec. 1972.

[5] G.S. FISHMAN "Estimating Sample Size in Computer Simulation Experiment" Manag. Sc., Vol. 18, No. 1, Sept. 1971, P. 24-38.

10 ACKNOWLEDGEMENTS

We are pleased to thank Mr.S.Rakkar for many helpful discussions and Telecom Australia for making available data from CUDN.

TABLE 1.
Program numbers and their processing times.

Program Number	Processing Time in ms	Program Number	Processing Time in ms
1	0.88	22	0.587
2	4.179	23	1.748
3	"	24	0.587
4	"	25	5.816
5	"	26	1.800
6	"	27	2.282
7	"	28	1.900
8	"	29	"
9	"	30	"
10	"	31	"
11	"	32	0.85
12	10.565	33	5.649
13	4.129	34	28.272
14	11.011	35	142.254
15	28.288	36	17.763
16	12.215	37	33.113
17	54.660	38	3.252
18	0.331	39	1.216
19	2.351	40	12.730
20	0.331	41	62.188
21	22.495	42	10.362

TABLE 2.
1972 input traffic rate for input messages.

Input Message	1972 Input Traffic Rate (Mean Arrival Rate) Messages/Second
LINK I/P A	5.99
LINK I/P B, C, D	0.75
LINK O/P A	5.99
LO SPEED I/P	0.195
LINK O/P B, C, D	0.729
LO SPEED O/P	0.338
LOCAL I/P A	0.650
LOCAL I/P D, C	0.142
LOCAL O/P A	0.650
LOCAL O/P D, C	0.142

TABLE 3.
Switching Centre Model.
The programs with priority 1, 26, 32 which are not listed in the table are used for interrupts and pollings.

Type of Message	Input Message	Required Program Numbers	
1	LINK I/P A	2, 18, 22	
2	LINK I/P B,C,D	3, 12, 19, 23, 37	
3	LINK O/P A	4, 20, 24	
4	LO SPEED I/P	5, 14, 27, 33, 34, 35, 41	
5	LINK O/P B,C,D	6, 15, 21, 25	OUTGOING MESSAGES
6	LO SPEED O/P	7, 17, 36, 40, 42	
7	LOCAL I/P A	8, 28	
8	LOCAL I/P D, C	9, 13, 30, 39	
9	LOCAL O/P A	10, 29	
10	LOCAL O/P D, C	11, 16, 31	

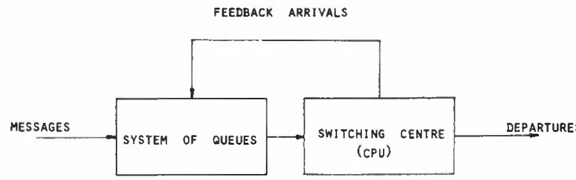


Fig. 1. Server facility model

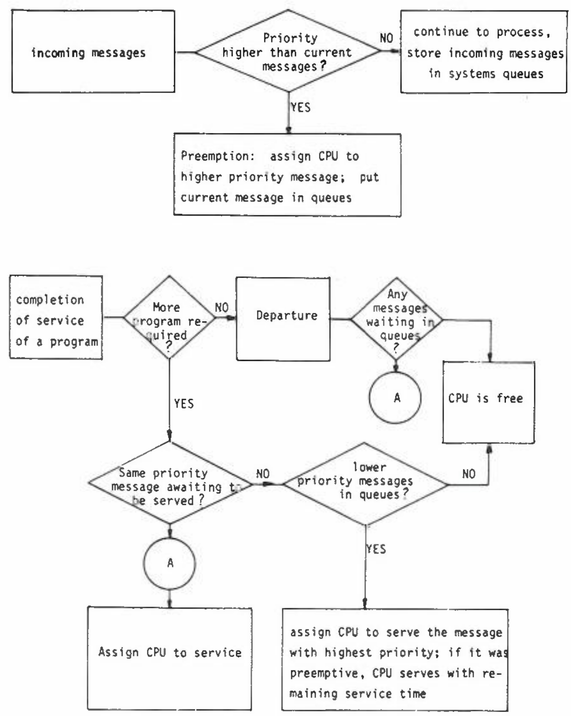


Fig. 2. Scheduling algorithm of switching centre.

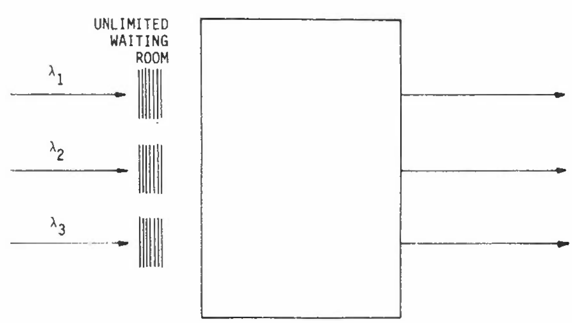


Fig. 3. Single server with various Erlangian time.

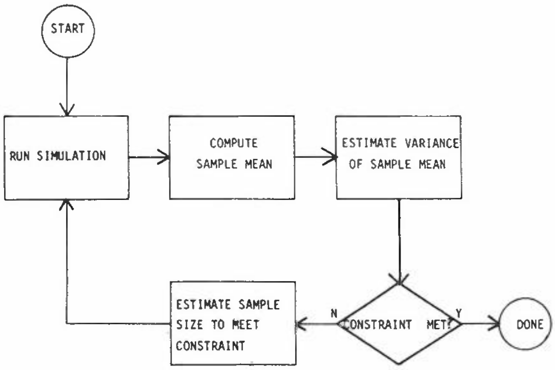


Fig. 4. Algorithm to estimate the sample size.

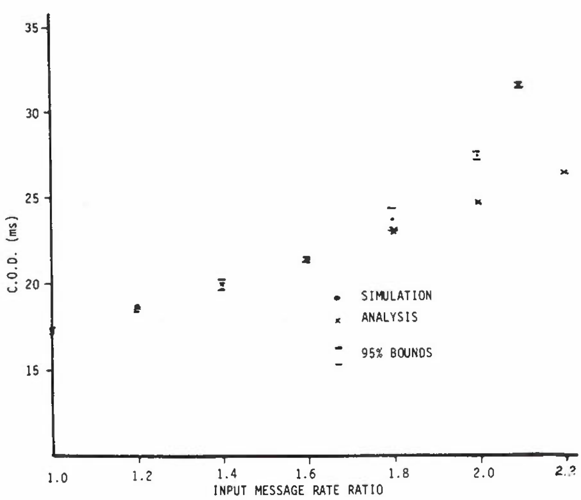


Fig. 5. C.O.D. of link A message vs input message rate ratio (where 1 equals the 1972 traffic rate).

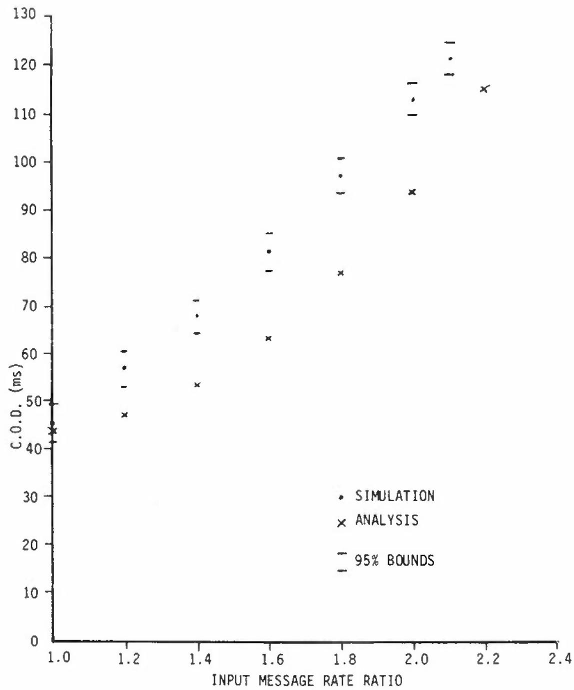


Fig. 6. C.O.D. of local priority message A vs input message rate ratio (where 1 equals the 1972 input traffic rate).

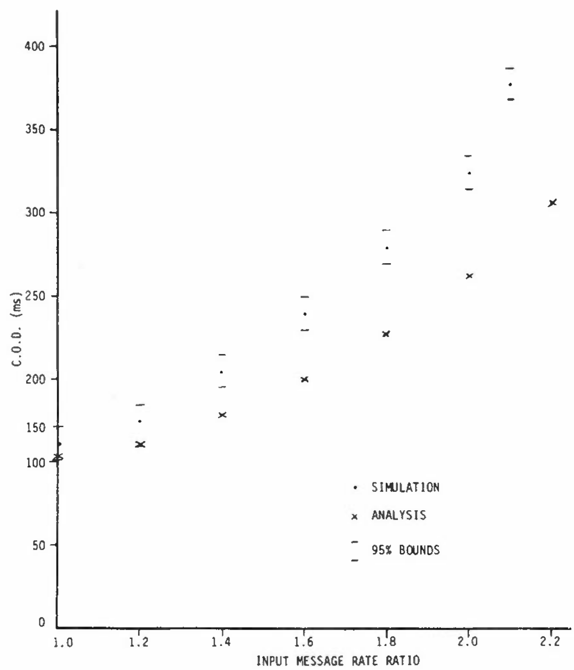


Fig. 7. C.O.D. of link B, C, D message vs input message rate ratio (where 1 equals the 1972 input traffic rate).

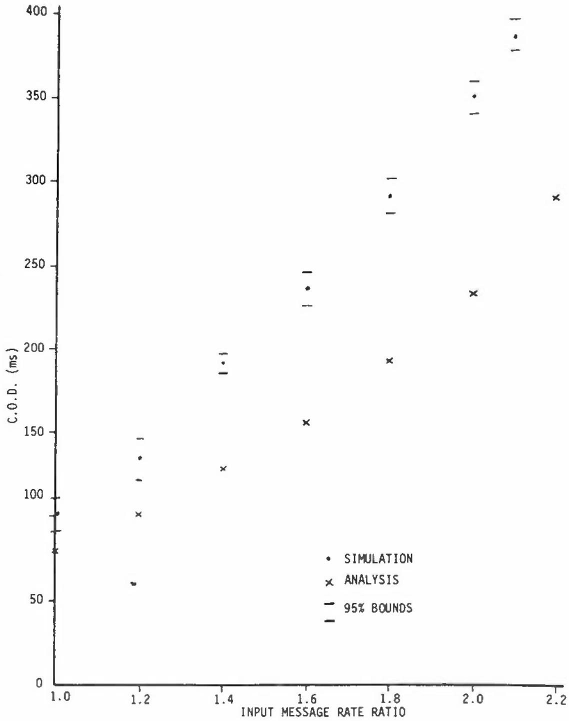


Fig. 8. C.O.D. of local message C, D vs input message rate ratio (where 1 equals the 1972 input traffic rate).

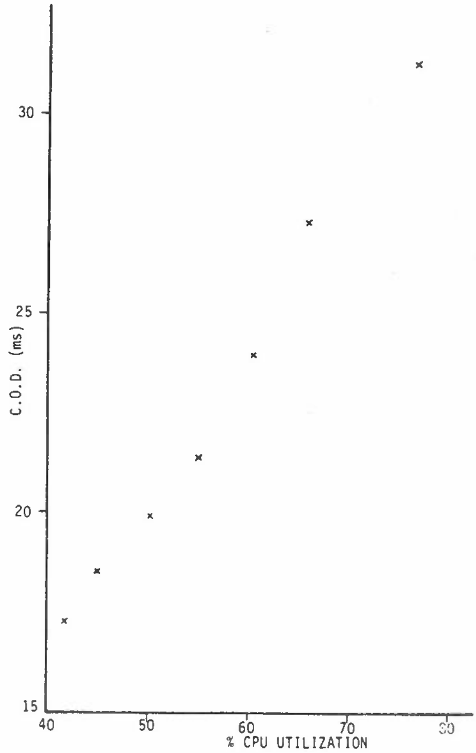


Fig. 10. C.O.D. of link A message vs CPU utilisation.

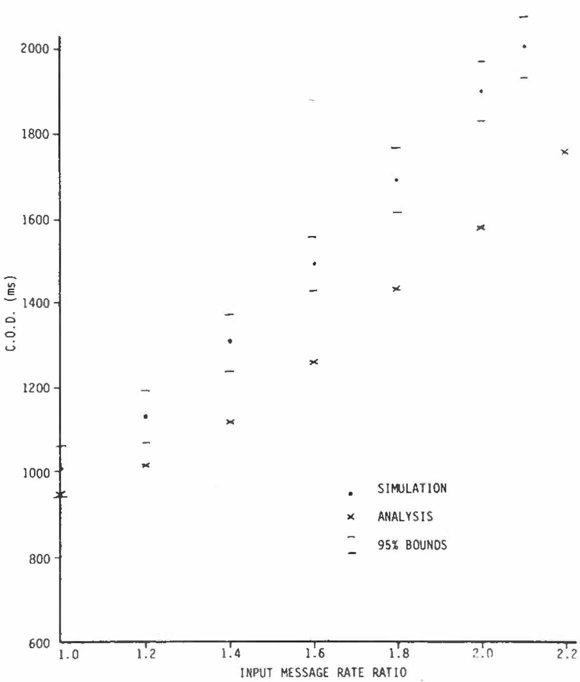


Fig. 9. C.O.D. of low speed input vs input message rate ratio (where 1 equals the 1972 input traffic rate).

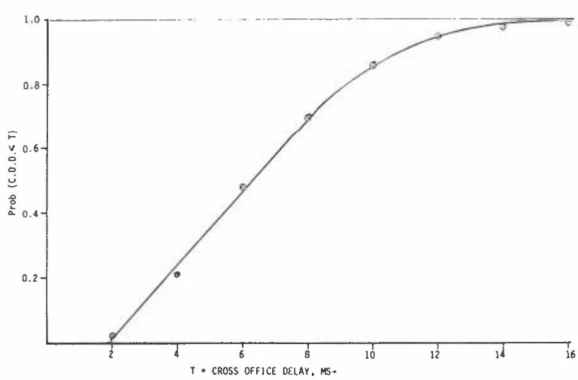


Fig. 11. Cumulative distribution for LINK input message A.

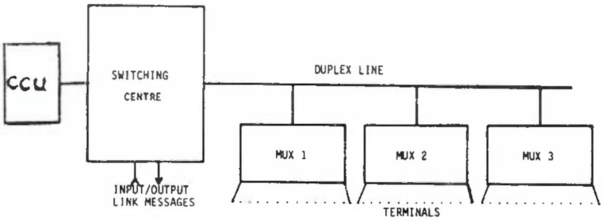


Fig. 12. Configuration of switching centre and customer computer unit (CCU).

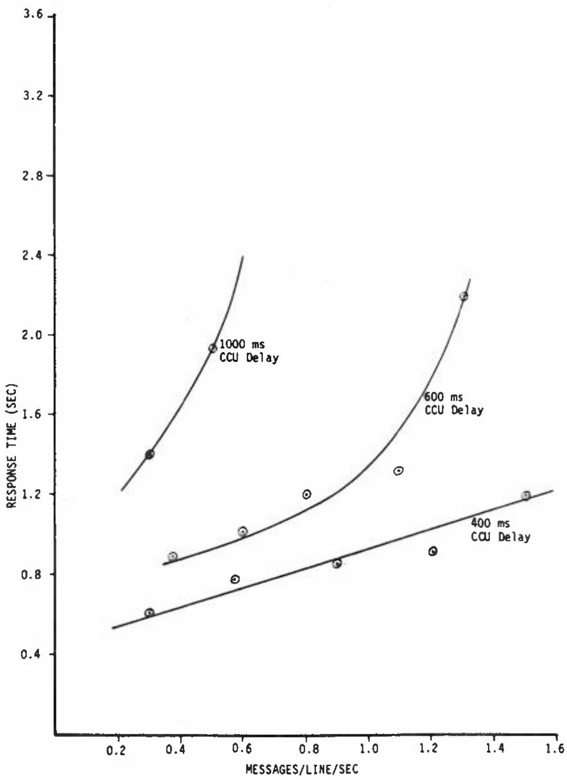


Fig. 13. Response time vs. throughput.

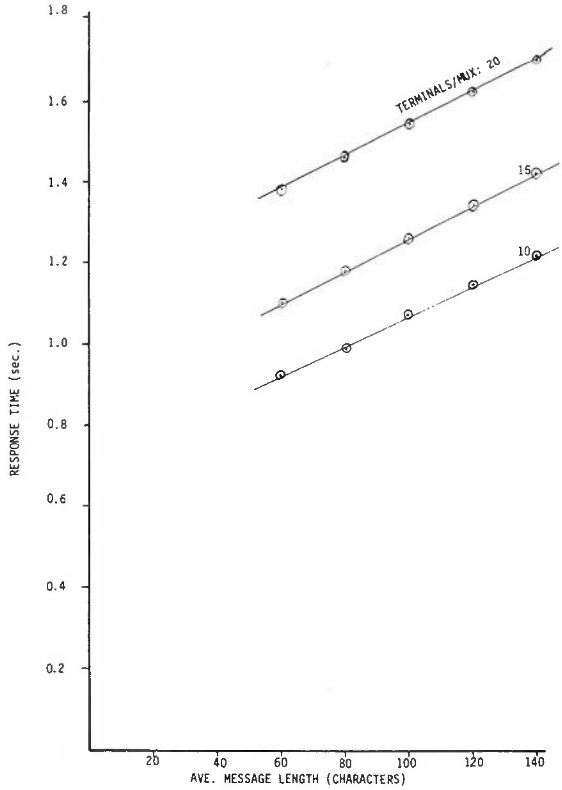


Fig. 14. Response time vs. average message length.

Discussion

J. RUBAS, Australia : *It is not clear from the figures 13 and 14 whether the small circles indicate results obtained by analysis or simulation as no confidence limits are shown. Could the authors explain, please?*

L.D. THUAN, Australia : *The Figures 13, 14 are obtained from simulation as discussed in Section 7. The confidence limits were also obtained but are not shown. The confidence limits are comparable to the sizes of the circles surrounding the plotted points.*

J. RUBAS, Australia : *In view of the great simplification of the real system in both the analytical and the simulation models used - and in the absence of live traffic measurements - on what grounds do the authors claim that the simulation results are more accurate than the analytical ones? Could the authors elucidate the reasons for the differences at higher input message ratios?*

L.D. THUAN, Australia : *We believe that the simplifications used in defining the simulation model are not significant with respect to the results presented. The main simplifications appear to be:*

(i) Some programs used in the real system have been excluded. These are in fact concerned with lower priority messages than those considered, and thus would be preempted by messages of the classes included. That is, they do not present a load in competition with those messages included in the

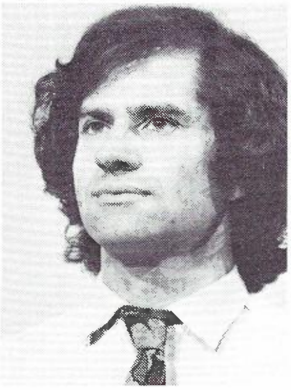
simulation whose performance is therefore not affected.

(ii) The load of housekeeping to put away tasks which are preempted and resume them is negligible compared with the main tasks.

(iii) Queues are of unlimited length.

To facilitate the application of existing queueing theory, the model had to be simplified. We assumed that messages at a particular priority level arrived independently at the same rate for each of the programs involved. This assumption is clearly not correct in the short term because preemption by high priority messages may cause the re-arrival of low priority messages to be bunched, i.e. these messages are no longer independent. This effect would be cumulative and of increasing severity at high utilization levels.

Our simulations could only be compared with real CUDN measurements under a limited number of conditions because (i) the measured data available was only for night time conditions because of the need to avoid interference with live traffic, (ii) it is not practical to apply experimentally a sufficient range of traffic (iii) the number of terminals is not under control and (iv) the CUDN systems applications programs have been changed since the data for which the simulations were carried out was gathered and thus it is not practical to gather more data relevant to the simulations. Of course the exercise could be repeated. Where it was possible to compare measured and simulated results there was good agreement.

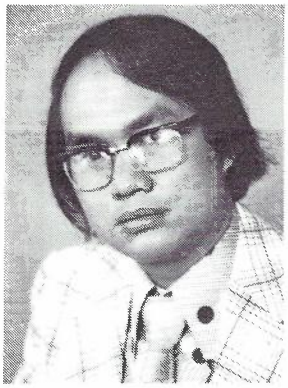


BIOGRAPHY

ROBERT EUGENE BOGNER was born in Melbourne in 1934. He graduated from the University of Adelaide in Electrical Engineering in 1956, while employed as a cadet engineer with the Postmaster General's Department. He researched speech signal processing for bandwidth reduction at the University of Adelaide while employed as an engineer with the P.M.G. Research Laboratories, gaining an M.E. for this in 1959 and continued work on signal processing, electroacoustics, and human factors in the P.M.G. Laboratories until the end of 1961. He was a lecturer and senior lecturer in the Electrical Engineering Department at the University of Queensland from then until 1967, with special interest in communication engineering, signal processing, electronics, including microwave modelling. From 1967 until 1973 he was a lecturer at the Imperial College of Science and Technology, University of London where he was active in communication engineering, signal processing, and speech communication. During this time he spent some months as a summer consultant at the Bell Telephone Laboratories, worked as a consultant for several companies and government establishments, and gained the Ph.D. degree and Diploma of Imperial College for work in phase processing of angle modulated signals.

Since August 1973 he has been Professor of Electrical Engineering and Chairman of the Department of Electrical Engineering in the University of Adelaide and in 1975 and 1976 was Dean of the Faculty of Engineering. His research interests continue to be primarily in the area of signal processing and communication, including human and social criteria in communication.

Professor Bogner is married with three children and enjoys among other things house repairing and squash.



BIOGRAPHY

LE DAC THUAN was born in Long Xuyen, Vietnam in 1949. He gained the Bachelor of Science Degree in (Mathematical) Sciences in 1972 and the Honours Degree in Electrical Engineering in 1973, both at the University of Adelaide. Through 1974 and 1975 he carried out research as a Master of Engineering candidate at the University of Adelaide on traffic studies of data networks and the present paper is based on part of this study. In July 1975 he joined Control Data in Melbourne as a Systems Analyst and his major activity there has been on systems design for a general wagering system (G.W.S.) for the Totalizer Agency Board in Victoria, Queensland and Natal (South Africa). Their system is currently being installed in TAB Queensland.

His main contributions for Systems Development is on the G.W.S "Stimulator" which can simulate several thousand "live" terminals. The Stimulator functions are to be used for Q.A. volume and flood testing, study of the data traffic vs number of processors, communications network loading etc.

CONCENT — An Aid to the Business Management of Telephone Networks

G. O'H. MARLOW

Telecom Australia, Adelaide, Australia

ABSTRACT

A method to measure telephone traffic levels in remote exchanges from a central location has been developed and the technique has been applied to a medium sized metropolitan network exceeding 200 000 subscribers. Six continuous seven day (24 hour) traffic studies have been carried out on this network over a 17 month period to establish the value of, and to observe any trends in, macro network traffic parameters. This paper discusses the reasons for these studies and outlines the results obtained to this stage. An interesting outcome has been the relative stability of the weekly load factor, which has important application for estimation of call earnings.

1. INTRODUCTION

Development of a telephone traffic measurement system known as the CENTOC (an acronym for CENtralised Traffic OCcupancy) system has enabled the simultaneous measurement of traffic levels in groups of circuits physically dispersed throughout a network (Ref.1). This method is based on the extension of measurement leads from selected circuit groups in remote exchanges, via cable pairs, to measuring and recording equipment at a central location. CENTOC has been used for four years in South Australia to measure and observe, on a fortnightly basis, the level of subscriber originating traffic generated in the Adelaide metropolitan network during the network busy hour. Other groups measured on this regular basis include 'back-bone' (final choice) routes within the metropolitan network and the more important trunk routes.

The lack of detailed and reliable information on several aspects of macro network behaviour, aspects which included daily and weekly traffic profiles and load factors for various types of traffic, and the realisation that these factors would not only vary over time but should be managed, led to CONCENT (derived from CONtinuous CENToc). By extending and modifying the facilities and equipment used for regular CENTOC measurements, a continuous seven day (24 hour) study of selected traffic groups in the Adelaide network became feasible.

The initial CONCENT study, in June 1974, had two immediate and general objectives, viz :

- To establish the value of important macro network traffic parameters and their inter-relationships;
- To provide an information base for recording any movements, either induced by external means or long term trends, in subscriber behaviour.

Because accurate information on macro network behaviour will become increasingly important for guiding the efficient operational management of the telecommunications network, five subsequent continuous seven day studies have been carried out at irregular intervals in the 17 months to December 1975. Although this period has not been of sufficient duration to establish any changing trends in subscriber behaviour, this paper will outline the results obtained to this stage.

All CONCENT measurements have been carried out in the Adelaide Telephone District (ATD), which is a seven-digit closed numbering area encompassing Adelaide, the capital city of the State of South Australia, and environs. Within the ATD, 240 000 exchange lines and 350 000 telephone stations (instruments) are connected to 64 local subscriber exchanges ranging in capacity from 100 to 15 000 lines. This area contains a population of approximately 900 000 (73% of that for South Australia) with average telephone densities of 27 exchange lines per 100 population and 0.6 exchange lines per residence. About 94% of ATD services have access to the Subscriber Trunk Dialling (STD) network and early in 1976 an estimated 75% of trunk calls originating within this area were dialled by subscribers.

This paper will discuss the principal objectives of the initial and of a continuing series of CONCENT measurements and the possible benefits flowing from these studies. The practical methodology is briefly explained and the results obtained to this stage are analysed in some detail. In particular, the relative stability of the weekly load factor is shown to have an important application for estimation of call earnings. Future development, utilising minicomputer control of the measurement system to provide regular or instantaneous reports, is outlined.

2. CONCENT OBJECTIVES

In order to achieve the general objectives outlined previously, it is necessary to firstly establish typical network traffic properties and secondly to be able to assess the extent of any future variation from initial observed values. Thus the detailed objectives of the CONCENT studies are to establish :-

- Daily and weekly traffic profiles;
- Time-dependent variation of idle capacity in the network;
- Daily and weekly load factors;
- Proportion of network traffic carried within specified time periods;
- A method to monitor network call earnings.

These five objectives are discussed in more detail.

2.1 DAILY AND WEEKLY TRAFFIC PROFILES

Very little information has been available in Australia regarding typical traffic levels apart from the traditional morning, afternoon and evening week day traffic study periods for each individual exchange of some two to three hours duration. At the micro level, information regarding traffic levels at off-peak periods has proved useful for specifying the optimum periods for the withdrawal of important sections of the transmission or switching network from service for maintenance reasons without undue risk of localised congestion developing in the network. Changes to traffic dispersion patterns beyond normal traffic measurement periods may lead to congestion on backbone routes and regular continuous observation, enabling early reaction to changing patterns, tends to obviate this problem. At the macro level there are several relationships such as the relative variations of subscriber originating traffic compared with trunk traffic, which are of interest to network planners.

2.2 TIME-DEPENDENT VARIATION OF NETWORK IDLE CAPACITY

The existing trunk tariff system in Australia has seven distance dependent tariff rates which may each be charged at day or night time scales, the latter being applicable between 1800-0800 hours local time. Within the next few years new tariff systems could be introduced with significantly increased scope for greater flexibility in charging for trunk calls. For the successful implementation of these measures, it will be important to estimate accurately the amount of trunk tariff stimulation to apply to the trunk network at any particular time. In this, there are two relevant variables, the price elasticity of demand for trunk calls for various categories of subscriber and the increased volume of traffic that it is planned to generate. (Note that the stimulation could be negative, as in the case of peak-load pricing). Methods are available to estimate the former using market research techniques and pilot studies etc., but CONCENT type measurements are necessary to assess and monitor the latter.

2.3 DAILY AND WEEKLY LOAD FACTORS

In this paper two load factors are defined as follows :

- Daily Load Factor : the ratio of the averaged daily traffic volume over 24 hours (Monday to Friday) to the time consistent busy hour (TCBH) traffic for the same (five day) period.
- Weekly Load Factor : the ratio of the weekly (seven day) traffic volume to the Monday to Friday TCBH traffic.

Information of this type is essential for planning studies that require estimation of annual calls or traffic when the only information available are infrequent measurements of TCBH traffic levels.

Load factors have considerable importance in another area; examination of the possibility of relating predicted growth rates or projections of busy hour traffic to that of annual calls. Circuit dimensioning and provisioning are based on maintaining specified grades of service during the time consistent busy hour for traffic levels averaged over the four consecutive busiest weeks of the year. However, targets for performance and growth are often expressed in terms of annual calls. Calls within a given time period are related to traffic by the average call hold time but an additional factor is involved in converting busy hour traffic to annual calls, viz. the load factor.

An objective of telecommunications management should be to increase the load factor, within the bounds of social acceptance, by carrying additional daily and weekly traffic volumes at the same level of busy hour traffic to obtain more efficient network utilisation.

2.4 PROPORTION OF TRAFFIC WITHIN SPECIFIED TIME PERIODS

While load factors enable the variation of busy hour traffic compared with total daily or weekly traffic volume to be observed, changes may occur in the daily time distribution pattern of a given volume of traffic through seasonal, cyclical or trend effects. By noting the percentage of daily traffic volume generated during a network's busiest three consecutive hours in the morning, afternoon and evening periods respectively, the more important changes to traffic patterns can be detected. Because tariffs on trunk routes in Australia are time as well as distance dependent, it is necessary for revenue estimation to establish the proportion of trunk traffic carried at night and day tariffs respectively. Thus the proportion of 24 hour traffic carried during the day rate period of 0800-1800 is required. Because erlanghour meters have oten been used in provincial trunk switching centres to record the traffic volume between 0900-1100 Monday to Friday, information is also required relating the TCBH traffic level to the average level for the time-switched period.

2.5 ESTIMATION OF NETWORK CALL EARNINGS

An important parameter for all business enterprises is accurate estimation of the current level of earnings. Because Telecom Australia must fund at least 50% of capital expenditure from internal sources, variations in call earnings compared with forecast levels could have significant repercussions on either general tariff levels or the proposed size of the capital investment programme for the ensuing year. Regular CONCENT measurements would enable fairly accurate estimates to be made of earnings trends from local and trunk calls and this information could be available to management within a few days of measurement.

3. METHODOLOGY

As stated earlier, the CONCENT Series of traffic measurements utilizes the CENTOC measuring and recording system. The details of this system are described in Ref. 1 and the general principles employed in traffic occupancy measurements in Australia are discussed in Ref. 2.

3.1 CONCENT

Because CONCENT measurements are continuous over 168 hours, the measuring equipment should operate automatically for the full study period to avoid high labour costs. This objective has been achieved by using Incremental Magnetic Tape Recorders (IMTRs) to output all measurement data on to magnetic tape. Four IMTRs are used which are each time switched sequentially to record for one six hourly period in every 24 hours. By this means all traffic information for the same six hourly period each day is contained on one reel of magnetic tape. Each tape is processed individually by computer to obtain half-hourly average readings for all traffic groups. Two hard copy outputs are available, 'CONCENT A' and 'CONCENT B' respectively. CONCENT A contains relatively raw data, listing the average traffic on each route at 48 half-hourly intervals commencing at midnight which is further sub-divided into three periods :-

- Average traffic Monday to Friday;
- Saturday traffic;
- Sunday traffic.

CONCENT B is a summary of results from elementary processing of the data from CONCENT A. For each traffic group it records :-

- Monday to Friday time consistent busy hour traffic and time of occurrence;
- Saturday and Sunday busy hour traffics and times;
- Average daily (24 hour) traffic volume, Monday to Friday (erlang-hours);
- Total weekly traffic volume over 7 days (erlang hours);
- Daily load factor Monday to Friday;
- Weekly load factor;
- Weekday afternoon and evening, Saturday and Sunday peak hour traffic levels as a percentage of average weekday morning busy hour traffic;
- Day rate traffic volume (0800-1800) as a percentage of average traffic volume over 24 hours (Monday to Friday);
- Busy hour traffic as a percentage of the average traffic level over the 0900-1100 period (Monday to Friday).

4. RESULTS

At the time of writing this paper six measurements, each of one week duration, had been undertaken in the CONCENT series. The first measurement was made in July 1974 with subsequent studies in May, July, August, September and December 1975.

In addition to the relatively slowly moving social, economic, cultural etc., changes which influence subscriber behaviour, there are several external factors which may have immediate impact. Chief among these, of course, is a change in the level of charges. Within the 17 month period under review, adjustments to telephone call charges were made on two separate occasions. These were :-

October 1 1974

- Local or unit call fee increased by 26% (to six cents);
- Most day trunk tariff rates increased, most night rates decreased;
- Fee for a trunk call booked with an operator which could have been dialled by the subscriber increased from 20 to 30 cents.

September 1 1975

- Local call fee increased by 50% (to nine cents);
- Day trunk tariff rates (0800-1800) increased by 50% for the four lowest tariffs (up to 325 Km) decreasing to zero change for calls over 645 Km;
- Night trunk tariff rates increased by 50% for the lowest tariff (less than 50 Km) with little other change;
- Fee for a trunk call booked with an operator which could have been dialled by the subscriber increased from 30 cents to 40 cents.

An additional factor that may be significant in its effect on telephone traffic has been the adoption of daylight saving in South Australia and some other States (but not all) by advancing local time by one hour between the end of October and the end of the following February each year.

This Section will analyse the results in two parts. Firstly, the results from a typical study will be considered and for this purpose, the July 1975 study will be reviewed. Secondly, the results from six studies extending over 17 months will be analysed in an attempt to detect any trends or changes occurring over time due either to policy and tariff changes, daylight saving or possible changes in customer calling patterns.

4.1 JULY 1975 STUDY

4.1.1 Traffic Profile and Busy Hour

In the CONCENT studies, the originating traffic from 34 subscriber exchanges were monitored, these exchanges all being within the ATD and exceeding 1000 connected lines in size. In July 1975, 28 or 82% had a Monday to Friday time consistent busy hour commencing either at 0900 (18 exchanges) or 0930 (10 exchanges). Only three exchanges, two of which serve the Central Business District (CBD) in the city, exhibited an afternoon busy hour, but the traffic level in each case did not exceed the corresponding morning peak by more than 1%. The combined ATD had a traffic TCBH which commenced at 0930 at a level of 8440 erlangs. This represents a calling rate of 0.039 erlang per exchange line. Afternoon and evening peaks were respectively 89% and 52% of morning levels. Table 1 shows the variation of peak hour traffic levels for afternoon and evening periods as percentages of morning peak levels for the four tandem switching areas into which the ATD is sub-divided. The Waymouth tandem area includes all four exchanges serving the CBD, as well as several exchanges in predominantly residential areas. Because of the markedly different calling patterns of these two groups, the tandem area has been further split into two sub-components.

Exchange Grouping	No. Xges Sampled	TCBH	Tfc (Erl)	% of Morning Peak Aft	% of Morning Peak Eve
ATD	34	0930	8440	89	52
Waymouth					
Tandem Area	9	0930	2880	94	38
City Xges	4	1100	1540	100	12
Metro Xges	5	0900	1420	86	65
Edwardstown					
Tandem Area	8	0900	2340	84	64
Nth Adelaide					
Tandem Area	11	0930	2540	90	54
Outer Metro					
Tandem Area	6	0900	700	91	62

TABLE 1. ATD Originating Traffic Details

The ATD busy hour for Saturday occurred at 1030 hours, an hour later than on weekdays, and at a traffic level of 45% of the weekday morning peak. This pattern was evident for each tandem area, although for the four city exchanges, in isolation, the traffic peak was only 15% of their average weekday levels. For Sunday, the busy hour for the ATD, and for each tandem area individually, occurred at 1100 hours when traffic was 38% of the weekday maximum. After excluding the four exchanges in the CBD, each tandem area recorded busy hour traffic between 40%-48% of weekday levels.

Figs. 1 to 4 illustrate several aspects of the variation of originating traffic over a one day or one week continuous period. In Fig. 1, the variation of ATD traffic over the full week of measurement is plotted. A similar pattern is evident from Monday to Friday with prominent morning, afternoon and evening peak periods followed by much lower levels and a less definite pattern at the weekend. Fig. 2 contrasts the characteristic traffic patterns exhibited by exchanges serving the CBD with those having a much stronger component of residential subscribers. The Waymouth tandem switching area of nine exchanges has again been divided, for illustrative purposes, into groupings of exchanges serving predominantly business (City) and residential (Metro) areas. The former grouping shows only two distinguishable daily busy periods with negligible traffic at evening or weekend periods, while the latter grouping has a markedly different profile featuring a strong evening peak period. In six exchanges, throughout the ATD, all strongly residential, the evening busy period traffic measured within 10% of afternoon maximum levels. For ATD originating traffic, Fig. 3 illustrates the relative levels, over a 24 hour period, of average Monday to Friday, Saturday and Sunday traffics respectively. Fig. 4, a normalised presentation of the data plotted in the previous figure, highlights the busy hour time shift on Saturday and Sunday and facilitates comparison of relative traffic levels throughout each period compared with the peak level.

All ATD backbone routes between tandem exchanges carried maximum traffic loads during the morning period and most routes had a TCHB commencing at 0930, the originating traffic busy hour. Most final choice routes between a local terminal exchange and its parent tandem exchange exhibited similar traffic profiles. However, ten terminal exchanges, all having a preponderance of residential services, had an evening busy period on their backbone route, with measured traffic reaching 60% above the level carried during the originating traffic busy hour. This apparent anomally is a reflection of a changed traffic dispersion during evening periods.

Trunk traffic originating from and terminating within the ATD exhibits a rather different profile compared with that for total originating traffic from subscriber exchanges. The afternoon and, in particular, the evening traffic peaks are relatively higher compared with the morning busy hour level and Saturday and Sunday busy periods occur during the evening. However, the principal point of interest is the contrast within the daily traffic profile depending whether the trunk call is dialled by the subscriber (STD) or is operator connected. Fig. 5 shows this variation over one week by splitting this traffic into its two components by method of connection. It is clear that STD peak traffic occurs during business hours while operator connected traffic has an evening peak period. In Fig. 6 the variation in STD traffic penetration, the percentage of STD to total trunk traffic, is plotted on a 24 hour time scale. It is clear from these graphs that outside of normal business hours, STD penetration falls considerably thus indicating the area on which management could concentrate to increase the overall level of penetration. For this and the next Fig., the penetration has not been plotted between mid-night and 0700 because the very small level of trunk traffic during this period makes the result meaningless.

Fig. 7 provides a relative comparison of trunk traffic penetration for traffic originating from and terminating within the ATD. During business hours between Monday and Friday, originating penetration is higher but the relationship is more variable at other times. The level of trunk traffic originating from the ATD expressed as a percentage of total subscriber originating traffic is shown in Fig. 8. The average level is about five percent during periods of reasonable traffic activity with peak levels close to eight percent.

4.1.2 Load Factors

Originating traffic load factors (see definition, Section 2) for the ATD and for each of its four tandem areas are given in Table 2. The Waymouth tandem area has again been split in two to demonstrate the significant differences between the City exchanges, serving the city, commercial and retail centre, and the other groupings.

Exchange Grouping	Daily L.F.	Weekly L.F.	% Day Rate	TCBH Tfc as % 0900-1100
ATD	9.4	55.4	78.3	102.7
Nth Adelaide Tandem Area	9.7	57.9	76.8	102.5
Edwardstown Tandem Area	9.3	56.5	74.4	104.1
Outer Metro Tandem Area	9.4	55.5	77.9	103.1
Waymouth Tandem Area	9.1	51.9	83.0	102.6
City Exchanges	8.4	44.4	91.4	106.7
Metro Exchanges	9.5	57.3	74.9	103.4

TABLE 2. Originating Traffic Load Factors

Excluding the group of city exchanges, there is considerable uniformity between these groupings for each of the four parameters listed in Table 2. The range of parameter values for the 34 individual exchanges, split into two groups, city exchanges (4) and metro exchanges (30), are given by Table 3.

	City Exchanges	Metro Exchanges
Daily Load Factor	7.5, 8.6	8.3, 11.0
Weekly Load Factor	38.2, 45.9	47.0, 68.9
% of Traffic 0800-1800	88.7, 94.9	68.3, 87.5
TCBH traffic as % of 0900-1100 average	106.2, 110.0	102.0, 108.9

TABLE 3. Range of Parameter Values

Fig. 9 shows the frequency distribution of weekly load factor for these 34 exchanges.

On final choice routes between tandem exchanges, the weekly load factor was generally in the range 20-40 while the percentage of daily traffic in the 0800-1800 interval ranged, in the main, between 75 and 80. That is, a percentage of traffic similar to that for subscriber originated traffic was carried during the traditional 'day' period although this type of traffic had a significantly different weekly (and daily) load factor.

On backbone routes between terminal exchanges and parent tandems, the range of values was, of course, far greater. The backbone route for one exchange, which is strongly residential in character and which has a significantly different traffic dispersion in the evening, recorded a weekly load factor approaching 120 with most of the traffic volume being carried after 1800 hours.

Overall, total trunk traffic generated in the ATD had quite similar load factors to those for total originating traffic but once again there were significant differences between the subscriber dialled and manual assistance components. The factors for both ATD originating and terminating traffic are given in Table 4.

	Daily L.F.	Weekly L.F.	% Day Rate
Originating Trunk Traffic	10.0	56.3	74.9
STD	9.2	50.6	80.5
Manual Assistance	10.2	61.5	61.0
Terminating Trunk Traffic	10.0	57.5	73.4
STD	9.6	54.1	76.9
Manual Assistance	11.0	66.4	65.2

TABLE 4. Trunk Traffic Load Factors

Compared with ATD originating traffic, the weekly load factor for trunk traffic (originating) is within 2%, and the ratio of traffic generated within the 'day' is within 5%.

4.2 THE SIX STUDIES - A COMPARISON

In this Section, the results from the six studies spanning the 17 month period July 1974 to December 1975 will be compared.

4.2.1 Traffic Profiles

For ATD originating traffic, the Monday to Friday time-consistent busy hour varied between 0900 (4 studies) and 0930 (2). For all except one study (August 1975), the Saturday busy hour lagged the weekday period by an half-hour or an hour. Similarly, the Sunday busy hour generally lagged the Saturday busy period by an hour. For three of the four tandem area groupings, the collective weekday busy hour was consistent for each of the six studies. Afternnon, evening, Saturday and Sunday traffic maxima for the ATD all maintained a relatively steady ratio to the morning peak level with the exception of the December 1975 study where the afternoon relative level was significantly lower than for the five earlier studies. This seemed to reflect a high Christmas peak in morning traffic rather than any decrease in afternoon traffic levels. A similar pattern was observed for each tandem area including the four city exchanges. Busy hour traffic details for the ATD including relative levels during other periods are given in Table 5.

Date of Study	M-F Busy Hour Tfc	M-F Busy Hour Time	% of M-F TCBH Traffic Aft	Eve	Sat	Sun
July 1974	8010	0900	90	50	43	40
May 1975	8670	0900	87	53	44	39
July 1975	8440	0930	89	52	45	38
Aug 1975	8650	0930	88	51	44	37
Sept 1975	8590	0900	87	54	45	37
Dec 1975	9250	0900	82	51	45	41

TABLE 5. ATD Traffic Levels

These figures do not however reveal the interesting and significant change in traffic profile that occurred in the December 1975 study. Fig. 10 is a normalised plot of average 24 hour Monday to Friday traffic variation for each of the six studies. The cross-hatched area is an envelope containing the traffic profiles for the first five studies with the December traffic profile plotted separately. This diagram clearly shows the half to one hour delay for the December evening traffic pattern compared with the earlier studies. This movement is probably due to a change of traffic profile in the summer months assisted by the adoption of daylight saving from November to February. Thus subscribers' evening traffic distribution is correlated to sun time rather than clock time. This contrasts with the morning traffic profile which is remarkably narrow and obviously dependent on clock time throughout all seasons of the year.

For total trunk traffic, the weekday busy hour fluctuated between 0930 and 1030 with consistent morning busy periods for STD traffic (commencing between 0930-1030) and evening busy periods for manual assistance trunk traffic (commencing between 1900-1930). However, on Saturday and Sunday, the daily peak traffic always occurred in the evening, commencing at 1800-1830 for Saturday and 1800-2030 for Sunday (the latter being for December 1975).

4.2.2 Load Factors

Daily and weekly load factors and the proportion of traffic occurring between 0800-1800 for ATD originating traffic for each study are shown by Table 6.

Date of Study	Daily Load Factor	Weekly Load Factor	Percent Traffic 0800-1800
July 1974	9.4	55.1	78.9
May 1975	9.4	55.4	77.3
July 1975	9.4	55.4	78.3
Aug 1975	9.3	54.8	78.7
Sept 1975	9.4	55.2	77.4
Dec 1975	9.6	56.7	76.2

TABLE 6. ATD Load Factors for Each Study

The values of daily and weekly load factor are remarkably uniform over the period, particularly when the result for the December 1975 study is excluded. Under this condition, the mean and standard deviation of the weekly load factor for the first five studies are 55.18 and 0.25 respectively.

Using a small sample distribution, the 99% confidence limits are 54.8, 55.5. It follows then, with high probability, that the December weekly load factor is significantly different compared with the five earlier studies. It is interesting to note, however, that the weekly load factor for the earlier five studies has a maximum/minimum ratio range less than 1.25% of the minimum while the busy hour traffic range is 8.3%. The comparative statistics when the December study is included are approximately 3.5% and 15.5% respectively. This uniformity over widely varying conditions has some important implications for revenue estimation. With regard to proportion of traffic occurring in the 0800-1800 period, there is again a broad degree of uniformity. Although the December study again diverges by the widest margin from the mean of the earlier five studies, this is not statistically significant at the 95% confidence level. No significant changes were observed in the volume of daily traffic carried within the morning, afternoon and evening busiest three hourly periods.

Load factors for trunk traffic, although not as consistent as for originating traffic, do show reasonable uniformity. Table 7 gives the load factor variations for trunk traffic originating within the ATD.

Date of Study	Daily Load Factor	Weekly Load Factor	Percent Traffic 0800-1800
July 1974	9.6	53.7	77.9
May 1975	10.3	61.8	73.2
July 1975	10.0	56.3	74.9
Aug 1975	10.3	59.0	73.7
Sept 1975	10.0	56.9	74.6
Dec 1975	10.1	57.4	71.8

TABLE 7. Trunk Traffic Load Factors

There are, however, significant differences within this general category between STD calls and operator connected (or manual assistance) trunk calls. These differences are clearly evident from Table 8, where the mean and standard deviation, taken over the six studies, have been calculated for each load factor and method of call establishment.

Type of Trunk Tfc	Daily LF Mean	Daily LF Std Dev	Weekly LF Mean	Weekly LF Std Dev	% Day Rate Mean	% Day Rate Std Dev
STD	9.1	0.2	50.8	1.8	80.5	1.3
Manual Assist	10.7	0.7	65.3	4.2	60.1	1.4
Total	10.0	0.3	57.5	2.7	74.3	2.1

TABLE 8. Comparison of Trunk Traffic Load Factors

From Table 8, it is evident that the daily and weekly load factors for STD traffic show a greater uniformity than for either manual assistance trunk traffic or for total trunk traffic. Partly, the comparatively large standard deviation for total trunk traffic load factors has been caused by the changing mix of STD and manual assistance traffic. Because of the different (lower) charge rates applying in Australia after 1800 hours, it is important for revenue estimation purposes to know the relative percentage of weekly traffic that is generated within each rate period. Not only are there significant differences in daily and weekly load factors for STD and manual assistance trunk traffic, but even more marked is the difference in the percentage of traffic carried at day and night rates. In broad terms, 20% of STD and 40% of operator trunk traffic is generated at night charging rates.

4.3 ESTIMATION OF CALL REVENUE

The stability of the weekly load factor for subscriber originating traffic over a wide range of traffic levels and under varying conditions has been a feature of these studies. By scheduling no more than, perhaps, two or three continuous seven day studies each year to observe for any long term trend in the load factor or for changes caused by a new tariff schedule, reasonably accurate observations of trends in call revenue earnings should be possilbe. Network busy hour traffic levels, measured on a weekly or fortnightly basis using the CENTOC system, would be averaged over a month and compared with the corresponding month or accumulated period for the previous year. Given the observed stability in the load factor, this method should provide an estimate of local call earnings for the current year accurate to within a few percentage points. This estimate would be a valuable aid to management.

STD call revenue earnings can be measured accurately by using either the CONCENT method or by using calibrated erlanghour meters (a current summing device) to measure traffic at each trunk charging rate within the central automatic trunk switching exchange. Estimates of earnings from trunk calls booked through the manual assistance operators is currently available on a monthly basis.

Hence CONCENT has provided the opportunity to economically and accurately estimate the level of total call revenue being earned within the ATD.

For a given load factor, three other variables could influence the rate of local call earnings - call hold time, the proportion of calls to non-revenue earning destinations, e.g., number information etc., and the level of congestion in the network. The first two parameters are derived from regular dispersion studies conducted at each terminal exchange over 1000 lines within the network, while estimates of congestion can be derived from several indicators. Thus, changes to these factors can be allowed for, if necessary, when estimating call earnings.

5. FUTURE DEVELOPMENT

There are two principal limitations preventing a full-time, continuous measurement system covering large regional areas at this time. Firstly, remote traffic groups are at present extended to the cental measuring point over physical circuits and the cost of expanding this method beyond metropolitan boundaries becomes prohibitive. However, a new system is being developed which will measure and transmit distant traffic information on up to eight traffic groups upon command from a central facility. This system, which will operate over carrier frequency derived circuits, uses telegraph channel bandwidth (120 Hz). Secondly, the cost of data processing becomes excessive. However, by using a mini-computer as the central controller, the data could be analysed, validated and corrected as it is received and average traffic levels or accumulated traffic volumes could be either recorded on magnetic tape for further processing or printed out at appropriate intervals.

When both of these developments have been implemented, it will be possible to monitor the complete network continuously and traffic congestion in any section of the network would be immediately identified.

6. SUMMARY

This paper has outlined the aims and objectives of a series of measurements using the CONCENT system and has briefly discussed the methodology adopted. The results obtained from an investigation covering some 34 local terminal exchanges serving over 200 000 exchange lines have been reviewed. An important outcome has been the stability of the weekly load factor for subscriber originating traffic over a wide range of traffic levels and conditions. This has opened the way to accurate estimation of local call earnings by regular measurement of network busy hour traffic levels. Future development and refinement of the system should lead to continuous monitoring and measurement of the network - an important network and business management tool.

7. ACKNOWLEDGEMENTS

The advice and assistance of Mr E Menkins, who contributed significantly to the CONCENT project, is gratefully acknowledged. Mr D Beck provided valuable assistance towards preparation of the graphs. The author also acknowledges the permission of Telecom Australia to present this paper.

8. REFERENCES

1. Marlow, G.O'H.: 'CENTOC - Towards Centralised Traffic Measurement'; Telecom. Journal of Aust., 1976, Vol. 26, No.1, pp 72-81.

2. Pratt, C.W.: 'Development and Application of Telephone Traffic Measuring Equipment (Part 1)'; Telecom. Journal of Aust., October 1973, Vol. 23, No.3, pp 205-209.

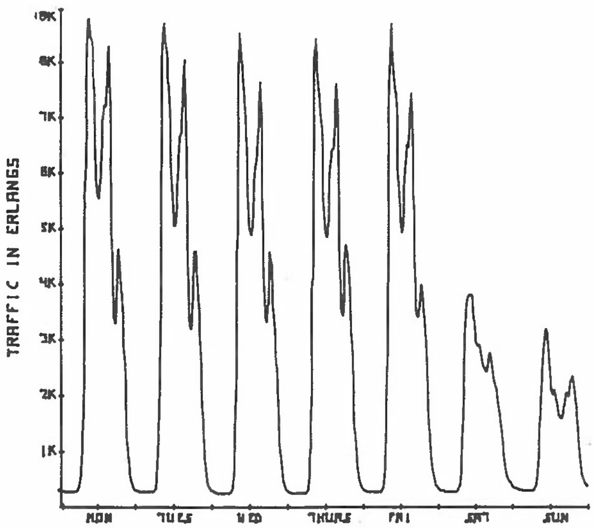


Fig. 1 Profile of ATD Originating Traffic Over One Week.

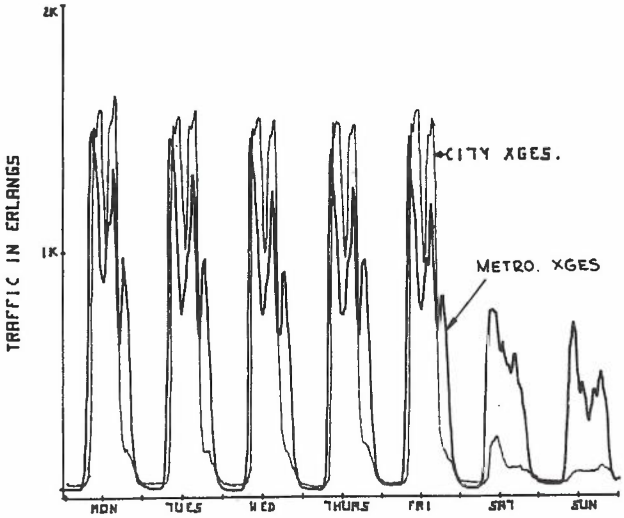


Fig. 2 Traffic Profile for Business and Residential Exchanges.

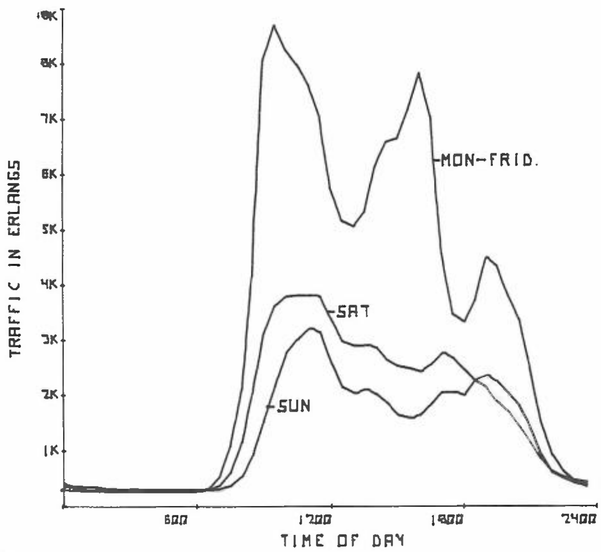


Fig. 3 Profiles of ATD Orig. Traffic over 24 hours.

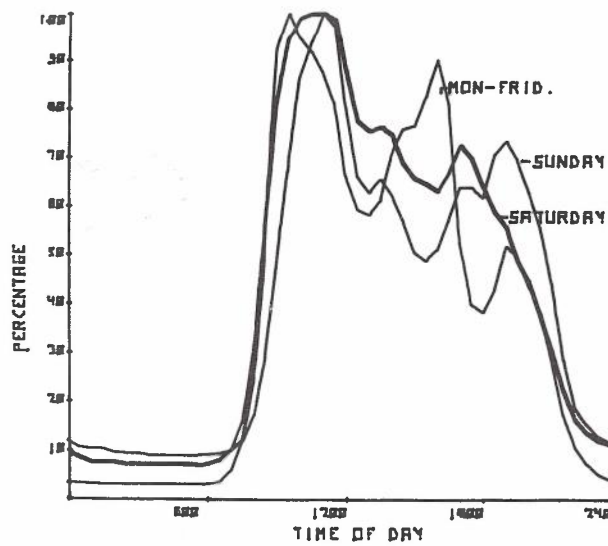


Fig. 4 Normalised Profiles of ATD Orig. Traffic

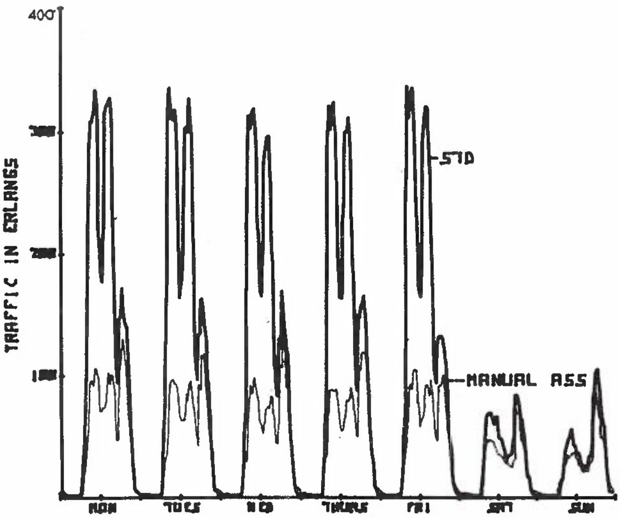


Fig. 5 Comparison of Trunk Traffic Profiles

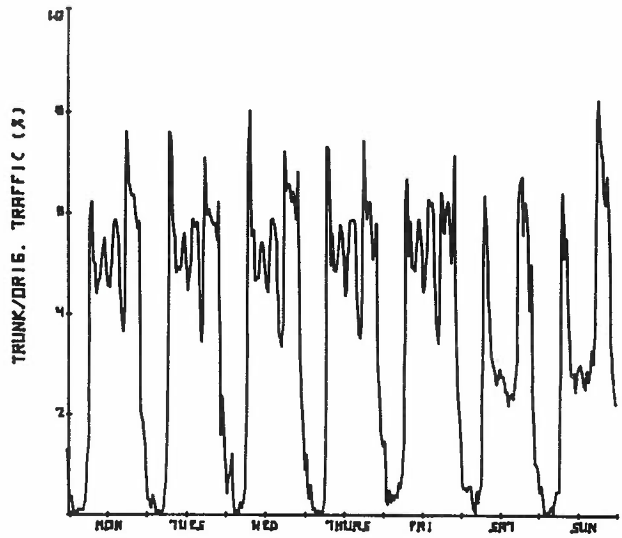


Fig. 8 Relationship of Trunk to Subscriber Orig. Traffic

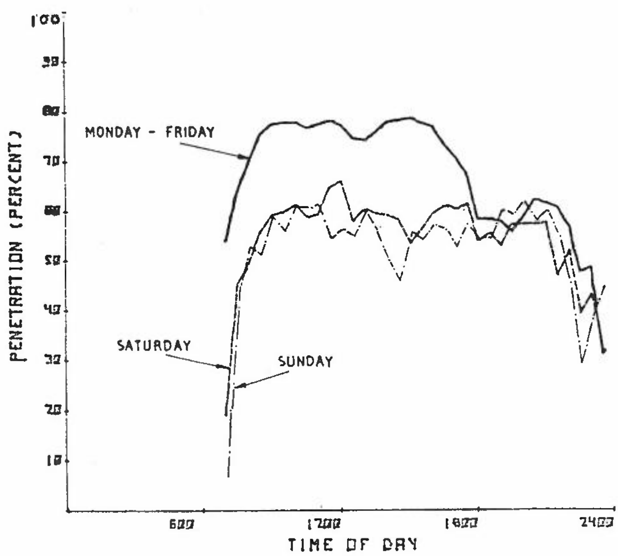


Fig. 6 STD Orig Trunk Traffic Penetration over 24 hours

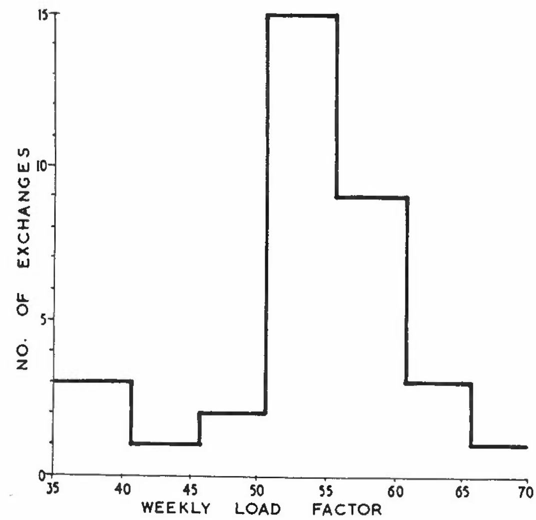


Fig. 9 Frequency Distribution of Weekly Load Factors

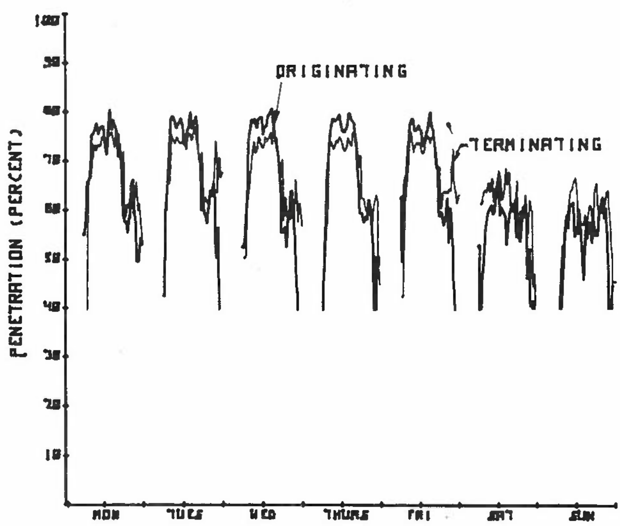


Fig. 7 STD Penetration Over One Week

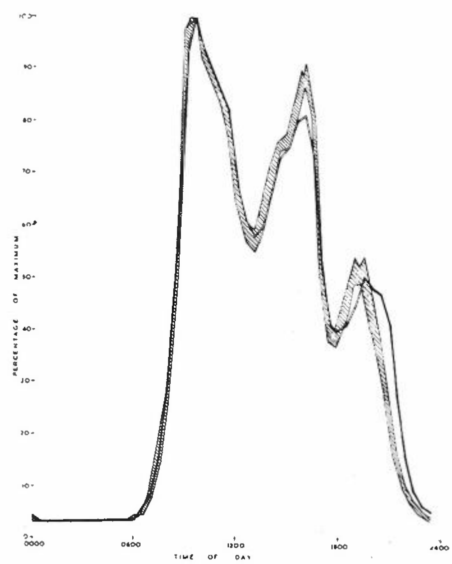


Fig. 10 Normalised ATD Orig Traffic for Six Studies.

Discussion

R. GREEN, Australia : The measuring method obviously permits the variance to mean ratio to be calculated. Did you calculate them and if so, can you say :-

(i) Was $^V/M$ = 1 at all exchanges at TCBH's.

(ii) What was the order and range of $^V/M$ of originating traffic outside the busy hours.

(iii) What was the order and range of $^V/M$ on the 'back-bone' or final choice routes.

G. MARLOW, Australia : We did calculate this ratio for time consistent busy hour (TCBH) traffic levels.

(i) For TCBH total originating traffic within each exchange, $^V/M$ ranged between 0.9-3.5. The values above about 1.5 occurred in 6 of 34 exchanges measured, and were due to very large changes in traffic levels during the TCBH. The TCBH was calculated on the busiest consecutive clock half hours and the first half hour for these 6 exchanges contained the very steep increases in traffic levels occurring shortly after 9.00 a.m.

(ii) These ratios, outside the busy hour, were not calculated.

(iii) During the TCBH, final choice routes generally exhibited the following ranges of variance/mean ratio :-

- terminal to tandem, 2.5-3.5
- tandem to terminal, about 2.0
- tandem to tandem, 2.5-3.5

Thank you for your question.

L. LEE, Canada : I have a problem to distinguish the originating from the terminating curves in Fig. 7 of your paper. Please help me by stating whether originating traffic is larger, or terminating traffic is larger, or sometimes one is larger than the other. Also please explain the reason behind this phenomenon. Thank you.

G. MARLOW, Australia : I apologise for the lack of clarity in Fig. 7. Originating penetration is higher from Monday to Friday, during business hours, and terminating penetration is higher on Saturday and Sunday. This variation can probably be attributed, in the main, to the influence of business services in Adelaide. After business hours and at weekends, terminating penetration is almost invariably higher than the originating level. Thank you for your question.

T. SUZUKI, Japan : Would you please comment on seasonal busy hour traffic variation in Australia and influence of busy season traffic on the load factor.

G. MARLOW, Australia : The seasonal traffic peak in Australia usually occurs immediately prior to Christmas but on a few routes the peak may be pre-Easter. The traffic levels at these seasonal peaks may be typically 5-10% above average busy hour levels during the year. For example, from Table 5 of the paper, it can be seen that the December 1975 busy hour traffic was 7.7% above the average for the four earlier studies in that year. Table 6 shows that only a relatively small variation to the daily and weekly load factors occurred at the December 1975 traffic peak compared with earlier studies. Thank you for your question.

L. GIMPELSON, Belgium : As the last paper in this session (by Graves and Pearson) and others have indicated, there is a close relation between traffic measurements and maintenance aids, since most of the measurements used for both tasks are the same. And certainly both contribute to that illusive grade-of-service concept. So, even though this is the Teletraffic Congress rather than ISS (Int. Switching Symposium) I'll ask why you chose to build a non-real-time system for traffic measurement reporting only, rather than combining it with maintenance requirements yielding a marginally more expensive quasi-real-time system which would cover traffic, grade-of-service and maintenance.

G. MARLOW, Australia : The answer to this question involves both historical and technical factors. From the technical viewpoint, it is important to note that measurements of telephone traffic in Australia are based on the traffic group, where the total current flowing through the traffic measuring equipment is proportional to the number of circuits in use. Unlike the method of individual circuit measurement, this technique is not generally suitable as a maintenance aid.

Historically, the CENTOC system of measurement developed as an extension to the method normally used in our exchanges. Instead of extending traffic groups by cable from the equipment being measured to the traffic measuring location within that exchange, selected groups were extended via junction cable to a central location within the city, due allowance being made during computer processing for the effect of the junction cable resistance. Although this system is not real-time, it did enable us to implement the concept quickly.

In order to extend the system throughout South Australia, a remote traffic recorder is being developed with a capacity of 128 groups which, combined with a mini-computer central controller, would enable real-time traffic monitoring throughout the network. Thank you for your question.

T. SUZUKI, Japan : Would you please comment on the following two questions.

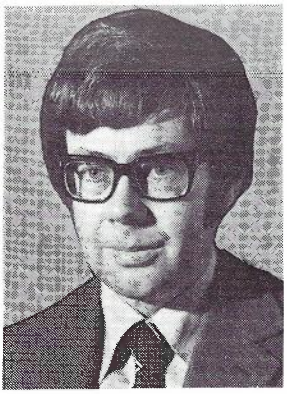
(i) Fig. 9 shows that several offices have a significantly different weekly load factor.

I think that use of the average weekly factor in estimating individual office earning and weekly traffic volume incurs significant error. How about this.

(ii) What influence did the call charge adjustment exert on the load factor and traffic volume.

G. MARLOW, Australia : My paper shows that the weekly load factor for subscriber originating traffic in the Adelaide metropolitan network has been very stable over a 17 month study period. This demonstrated stability should enable estimation of weekly network traffic volume and hence call earnings from regular measurements of busy hour traffic. The paper does not claim that this method would be accurate for the determination of traffic volume or earnings for a particular, individual exchange. The normal statistical variations which are observed for individual exchanges are averaged out over a network of exchanges.

With reference to your second question, the paper shows that the call charge adjustment had no discernible effect on the load factor. The fee increase did not appear to influence the level of originating traffic, but there was a significant swing towards usage of STD following increases in the manual call surcharge in 1974 and 1975. Thank you for your questions.



BIOGRAPHY

GEOFF MARLOW graduated from the University of Adelaide in 1964 as a Bachelor of Engineering after studying under a PMG cadetship. He worked in the Workshops Section for 3 years, including a short period at Headquarters concentrating on conversion of State Workshops to the manufacture of crossbar relays and relay sets. In 1968 he was transferred to Whyalla as a District Engineer where he was responsible for the maintenance of the East-West Microwave System within South Australia. In 1972 Mr. Marlow transferred to the Planning and Programming Branch where he worked for three years in the Traffic Engineering Section concentrating on problems of network management. He received a B.Ec. degree from the University of Adelaide in 1975 and is currently a Senior Engineer in the Fundamental Planning Section.

A Comparison of System and User Optimised Telephone Networks

R. J. HARRIS

Telecom Australia, Melbourne, Australia

ABSTRACT

In a recent paper (Ref. 1) the principles of System and User optimization were introduced and discussed for alternate routing telephone networks. (A System optimized network design is obtained by minimising the total cost of the network, subject to Origin to Destination grade of service standards. A User optimized network design is achieved by minimising the cost per erlang on chains used by each OD pair). An algorithm for determining these optimal network designs has been developed (Ref. 2), which is based upon a modification to a well known non-linear programming algorithm proposed by Wolfe (Ref. 3).

A mathematical model for dimensioning alternate routing networks developed by Berry (Ref. 6) has been used in conjunction with the optimizing algorithm to obtain network designs based on the two principles. The purpose of this paper is to compare the two different network designs obtained by applying these optimizing principles to the Adelaide Telephone Network.

1. INTRODUCTION

This paper will be concerned with two optimizing principles which were first discussed for telephone networks in Reference 1. These two principles are known as System and User optimization respectively. In the earlier paper, these principles were applied to a small network which consisted of five links; in the present paper these principles will be applied to the Adelaide Telephone Network using Berry's network Dimensioning model, and a comparison will be made of the two different networks produced.

There were two main factors which influenced the study of User optimization for telephone networks. Firstly, this form of optimization was new to telephone networks and represents an alternative to the principle of System optimization which is the ultimate aim of telephone network planners. Secondly, experience with this type of optimization in the context of road traffic theory had shown that the difference in costs between a System and a User optimized network was relatively small and furthermore, the speed with which the User optimized solution can be computed using Berry's dimensioning formulae is significantly greater than the computation speed required for System optimization. Thus it was felt that the User optimization procedure could be used to move rapidly to a network cost which was very close to the true System optimal solution. After reaching this point it would then be possible to step quickly to the System solution and hence to make a considerable saving in the overall computation time required to achieve the minimum cost network.

As mentioned above the optimization was performed using Berry's dimensioning formulae. The original formulae proposed by Berry have been recently upgraded to give more accurate results through the introduction of a new parameter "p" (Ref. 5). If this parameter is set to zero, then the formulae reduce to their original form. In this study, the old and the new versions have been used for computing circuit values required in the optimization process. It is not the intention of this paper to compare the two models as this has already been reported (c.f. Ref. 4). (Furthermore, different sets of data were used with the two versions of the dimensioning formulae).

In Section 2 of this paper, some theoretical results will be presented which have formed the basis for calculating a User optimized telephone network. (These results will be derived using the upgraded formulae - in each case the results simplify if a corresponding result is required for the original dimensioning formulae). In Section 3, there will be a brief discussion of the User optimized solution to the Adelaide Network, and in the following section a comparison will be made of the significant similarities and differences between the System and User optimized solutions.

2. THEORETICAL RESULTS USING BERRY'S DIMENSIONING FORMULAE

2.1-SYMBOLS AND DEFINITIONS

Consider a telephone network T consisting of a set N of exchanges, and a set L of u directed links. In this network we specify a set of n Origin to Destination pairs, (abbreviated to "OD pairs") and denote them by O^k-D^k, $k=1,\ldots,n$, and a set of m(k) allowable chains denoted by m_j^k for $j=1,\ldots,m(k)$. (A chain may be described as a sequence of links which form a route between an exchange which originates traffic and an exchange which terminates this traffic). Let t^k be the offered pure chance traffic destined for D^k from O^k. The amount of traffic carried on a particular sequence of links forming a chain between a pair of exchanges is called a <u>chain flow</u> and is denoted by h_j^k. The total carried traffic f_i on a link i is obtained by adding together the chain flows which use this link, i.e.

$$f_i = \sum_k \sum_j a_{ij}^k h_j^k \qquad i=1,\ldots,u. \tag{1}$$

where

$$a_{ij}^k = \begin{cases} 1 & \text{if link } i \text{ is on chain } m_j^k. \\ 0 & \text{otherwise.} \end{cases}$$

In addition to the above notation, the following symbols will be used :

- i symbol used to indicate an arbitrary link,
- j symbol used to indicate an arbitrary chain,
- k symbol used to indicate an arbitrary OD pair,
- M_i the mean offered traffic to link i,
- V_i the variance of the traffic offered to link i
- v_i the variance of the traffic overflowing from link i,
- C the total cost of the network
- B^k the traffic congestion between the k-th OD pair,
- n_i the number of circuits required on link i (regarded as a continuous variable for theoretical convenience).
- F^k the carried traffic on a direct route m_1^k.

In an earlier paper which discussed the concepts of System and User optimization for telephone networks (see Ref. 1), the notions of marginal chain cost and chain cost per unit flow were introduced in order to specify optimality conditions for these concepts. For completeness these notions and definitions of System and User optimization are summarised below :

Definition 1

If C(h) is the total cost of the network then the marginal chain cost of chain m_j^k is defined as

$$D_j^k(h) = \frac{\partial C(h)}{\partial h_j^k} \qquad (2)$$

Definition 2

For alternate routing networks, the cost per unit flow on chain m_j^k is

$$\bar{C}_j^k = \sum_i a_{ij}^k \, \bar{c}_i \qquad (3)$$

where

$$\bar{c}_i = \begin{cases} \dfrac{c_i n_i}{f_i} & \text{if } f_i \neq 0 \\ \lim\limits_{f_i \to 0} \dfrac{c_i n_i}{f_i} & \text{if } f_i = 0 \end{cases} \qquad (4)$$

and c_i = the cost per circuit on link i.

Definition 3

A network T is said to be "System optimized" if it is designed in such a way that the total network cost C(h) is a minimum, subject to OD pair grade of service constraints. A chain flow pattern is said to satisfy the conditions of System optimality if for each OD pair k, there is an ordering $p_1, p_2, \ldots, p_s, p_{s+1}, \ldots, p_{m(k)}$ of the chains joining O^k to D^k such that the following condition holds :

$$\lambda_k = D_{p_1}^k = \ldots = D_{p_s}^k \leq D_{p_{s+1}}^k \leq \ldots \leq D_{p_{m(k)}}^k \qquad (5)$$

where

$$h_{p_r}^k > 0 \quad \text{for } r=1,\ldots,s$$

$$h_{p_r}^k = 0 \quad \text{for } r=s+1,\ldots,m(k)$$

Definition 4

A chain flow pattern h which satisfies the OD grade of service constraints is said to satisfy the equilibrium conditions for a User optimized network T if, for each OD pair k, there is an ordering of the chains $p_1, \ldots, p_s, p_{s+1}, \ldots, p_{m(k)}$, joining O^k to D^k such that

$$\mu_k = \bar{C}_{p_1}^k = \ldots = \bar{C}_{p_s}^k \leq \bar{C}_{p_{s+1}}^k \leq \ldots \leq \bar{C}_{p_{m(k)}}^k \qquad (6)$$

where

$$h_{p_r}^k > 0 \quad \text{for } r=1,\ldots,s$$

$$h_{p_r}^k = 0 \quad \text{for } r=s+1,\ldots,m(k)$$

2.2 BERRY'S MODEL (OUTLINE)

The aim of Berry's model (Refs. 5 & 6) is to produce a minimum cost network (System optimized) subject to constraints on the chain flow variables which specify the Origin-Destination requirements. The model is formulated as a mathematical programming problem, and various techniques based on non-linear programming algorithms may be used to determine the optimal network. For the purposes of this paper, the circuit dimensioning formulae from Berry's model have been used with the modified non-linear programming model (Ref. 2), to study the two types of optimal network design.

Starting with a prescribed chain flow pattern on the network, the mathematical model determines the mean (M_i) and the Variance (V_i) of combined traffic streams offered to each link i, and calculates the equivalent random traffic (A_i) which would produce an overflow traffic distribution from a full availability link with moments M_i and V_i. The total traffic carried on each link (f_i) is determined from the sum of chain flows using that link (see equation (1)), from which the number of circuits (n_i) is calculated using equations (7) and (8).

$$n_i = f_i + A_i \left[\frac{(M_i - f_i)}{(M_i - f_i)(M_i - f_i - 1) + v_i} - \frac{M_i}{M_i^2 - M_i + V_i} \right] \qquad (7)$$

where the variance of the overflow traffic from link i is given by

$$v_i = \frac{1}{6}(M_i - f_i)\left(3 - M_i + f_i + \sqrt{(M_i - f_i - 3)^2 + 12A_i(1 - f_i/M_i)^p}\right) \qquad (8)$$

where p is an assigned parameter which takes on a value of approximately 0.1.

(The value of p has been determined from simulation trials and varies according to whether link is on a first, second or third choice route). A_i represents the equivalent pure chance traffic and is computed from the well known approximation by Rapp (Ref. 7).

$$A_i = V_i + 3\frac{V_i}{M_i}\left(\frac{V_i}{M_i} - 1\right) \qquad (9)$$

2.3 USER OPTIMIZATION THEORY APPLIED TO BERRY'S MODEL

Equation (4) from Definition 2 states that the link cost per unit flow (i.e. cost per unit erlang) is obtained from dividing the total cost of the link ($c_i n_i$) by the carried traffic on this link (f_i) - provided that this traffic is non zero. If the carried traffic is zero, then the link cost per unit flow is given by the limit :

$$\bar{c}_i = \lim_{f_i \to 0} \frac{c_i n_i}{f_i} \qquad (10)$$

In order to perform the User optimization with Berry's model it is necessary to compute the above limit.

Result 1

For a general link i, the limiting cost per unit flow is:

$$\bar{c}_i = c_i \left[1 + \frac{A_i M_i^2}{(M_i^2 - M_i + V_i)^2} \left\{ \frac{5V_i - 3M_i + M_i^2 + A_i p}{6V_i - 3M_i + M_i^2} \right\} \right] \qquad (11)$$

Proof: By rearranging equation (7) the required limit

$$\lim_{f_i \to 0} \frac{c_i}{f_i} \left[f_i + \frac{A_i(f_i(M_i^2 - M_i f_i - V_i) + M_i(V_i - v_i))}{(M_i^2 - M_i + V_i)((M_i - f_i)^2 - (M_i - f_i) + v_i)} \right]$$

consider the limit

$$\lim_{f_i \to 0} \frac{V_i - v_i}{f_i}$$

Using equation (9), V_i can be written as

$$V_i = \frac{1}{6} M_i\left(3 - M_i + \sqrt{(M_i - 3)^2 + 12A_i}\right)$$

hence

$$V_i - v_i = \frac{1}{6}\left\{(3 - 2M_i)f_i + f_i^2 + \sqrt{(M_i - 3)^2 M_i^2 + 12A_i M_i^2} - \sqrt{(M_i - f_i)^2 (M_i - f_i - 3)^2 + 12A_i(1 - f_i/M_i)^p (M_i - f_i)^2}\right\}$$

Dividing by f_i and taking the limit as $f_i \to 0$ gives

$$\lim_{f_i \to 0} \frac{V_i - v_i}{f_i} = \frac{1}{6}\left\{(3-2M_i) + \frac{(2M_i-3)(M_i-3) + 6A_i(2+p)}{\sqrt{(M_i-3)^2 + 12A_i}}\right\}$$

Thus

$$\lim_{f_i \to 0} \frac{c_i n_i}{f_i} = c_i\left[1 + \frac{A_i}{(M_i^2 - M_i + V_i)^2}\left\{M_i^2 - V_i + \frac{M_i}{6}\left[(3-2M_i) + \frac{(M_i-3)(2M_i-3) + 6A_i(2+p)}{\sqrt{(M_i-3)^2 + 12A_i}}\right]\right\}\right]$$

(since $\lim_{f_i \to 0} v_i = V_i$).

Substitution of equation (9) into the above expression and simplifying, the following result is obtained :

$$\lim_{f_i \to 0} \frac{c_i n_i}{f_i} = c_i\left[1 + \frac{A_i M_i^2}{(M_i^2 - M_i + V_i)^2}\left\{\frac{5V_i - 3M_i + M_i^2 + A_i p}{6V_i - 3M_i + M_i^2}\right\}\right]$$

Result 1 may be simplified further if the traffic is offered to a direct route (first choice route), as such routes are assumed to be offered Pure Chance traffic. Consequently, $A_i = M_i = V_i = t^k$ where t^k is the pure chance offered traffic to OD pair k on direct route k. Thus we obtain by direct substitution into equation (11):

Result 2

For direct route k, the limiting cost per unit flow is given by :

$$c_k\left[1 + \frac{(t^k + 2 + p)}{t^k(t^k + 3)}\right] \qquad (12)$$

In order to try and predict the form of the User optimizing chain flow pattern, two results have been obtained which examine the properties of the cost per unit flow functions. Result 3 shows that the magnitude of the number of circuits per unit flow is always greater than unity, and Result 4 demonstrates that the cost per unit flow function on a direct route increases with an increase in carried traffic, (offered traffic being held constant). From these two results it is possible to show (Result 5) that if the magnitude of the sum of the costs per circuit for links on second and subsequent choice chains is always greater than the largest value of the cost per unit flow on the first choice chain of a given OD pair, then a User optimized solution will have all flow on the first choice chain of this OD pair (see (17)).

Result 3

The cost per unit flow on a link is strictly greater than the cost per circuit. Equivalently,

$$\frac{n_i}{f_i} > 1 \qquad \text{for } f_i > 0$$

and

$$\lim_{f_i \to 0} \frac{n_i}{f_i} > 1 \qquad \text{if } f_i = 0 \qquad (13)$$

Proof: (i) For positive flow on link i the number of circuits is always greater than the flow since each term in equation (7) is positive.

(ii) For f_i=0, the limit given by equation (11) is always positive and greater than unity since the second term of this expression is positive. (This follows because $A_i \geq V_i \geq M_i > 0$).

Result 4

The cost per unit flow is a strictly increasing function of the flow for direct routes when the offered traffic is fixed.

Proof (Outline): It is necessary to show that the derivative of the cost per unit flow function for direct routes is positive.

Let $A_i = M_i = V_i = t$ and $f_i = F$, then substitution into equation (7) yields the following simplified formula for the number of circuits on a direct route :

$$n = F - 1 + t\left[\frac{(t-F)}{(t-F)^2 - (t-F) + v}\right] \qquad (14)$$

$$\text{where } v = \frac{1}{6}(t-F)(3 - t+F + \sqrt{(t-F-3)^2 + 12t(1-F/t)^p}) \qquad (15)$$

is the variance of the traffic overflowing from this route. For convenience the square root term above may be defined as X, i.e.

$$X = \sqrt{(t-F-3)^2 + 12t(1-F/t)^p}$$

Rearrangement of equation (14) yields

$$\frac{n}{F} = \left[1 + \frac{12(t+2+2F+p)}{(5(t-F)-3+X)(t+5F+3+X)}\right] \qquad (16)$$

It is now necessary to differentiate the above equation with respect to F and to determine whether the resulting expression is strictly positive. This computation has been performed and the derivative is strictly positive. (The details of this computation have been omitted as the expressions are rather unwieldy).

Result 5

In a telephone network, if condition (17) holds for an OD pair k, then a User optimized flow pattern for this network will have all flow on the first choice chain m_1^k.

$$c_k \beta_k < \sum_i a_{ij}^k c_i \qquad \text{for } j=2,\ldots,m(k) \qquad (17)$$

where c_k = Cost per circuit on direct route k,

c_i = Cost per circuit of link i on alternate route m_j^k.

$\beta_k = \left.\frac{n_k}{F^k}\right]_{F^k}$ = Maximum feasible flow on direct link k.

Proof: For first choice chains, the cost per unit flow for the direct route is an increasing function of the flow. Thus the cost per unit flow will have its maximum value when F^k is assigned its maximum feasible value, i.e.

$$F^k = t^k(1-B^k)$$

The cost per unit flow on the link of an alternate route is strictly greater than the cost per circuit of that link and hence

$$\bar{C}_j^k = \sum_i a_{ij}^k \bar{c}_i > \sum_i a_{ij}^k c_i \qquad \text{for } j=2,\ldots,m(k)$$

Now, if $\sum_i a_j^k C_i$ is greater than the maximum value of the cost per unit flow on the first choice chain, for j=2, ...,m(k) then

$$\bar{C}_j^k > \bar{C}_1^k \qquad \text{for } j=2,\ldots,m(k)$$

Thus from Definition 4 in any User optimized flow pattern it follows that there cannot be any flow on chains $m_2^k, m_3^k, \ldots, m_{m(k)}^k$.

---

3. SYSTEM AND USER OPTIMIZATION OF THE ADELAIDE NETWORK

3.1 SYSTEM OPTIMIZATION

In this study, two different sets of cost and traffic data were used with the two versions of Berry's dimensioning formulae. Data for the first version of the model specified 1141 OD pairs which connected 37 exchanges. Four tandem exchanges were provided for the alternate routes. Data for the second version of the model specified 1892 OD pairs which connected 44 exchanges, and once again four tandem exchanges were used for directing traffic on the alternate routes. The network uses cross-bar group selectors which provide traffic with a maximum of three possible choices.

In both optimizations, the starting point for iterations was "all flow on the viable direct routes and all remaining traffics assigned to second choice routes if no direct route is supplied". Iterations were terminated after approximately 15 minutes of CP time on a CDC Cyber 74 computer. In this time, approximately 500 iterations of the algorithm described in Reference 2 were required to achieve a chain flow pattern which was accurate to within one or two percent of the true System optimal solution. In practice, there are several reasons for not attempting to obtain a more precise solution. Firstly, a System optimized network is very sensitive to traffic overload, and this is considered to be undesirable from a practical viewpoint. Secondly, the formulae use circuit values which are continuous variables (for theoretical convenience) and they must be rounded to integers at the completion of the optimization process. Experience has shown that very little change in the overall integer solution cost is made by continuing to refine the accuracy of the solution. Finally, the convergence of the optimization process is particularly slow near the minimum and further application of the algorithm would require a considerable computational effort which would not yield any advantages in terms of the integer solution.

Table 1 summarises the network costs for every twentieth iterate for the two optimizations. In Table 2 a selection of OD pair chain flows and their corresponding marginal chain costs is presented for which Definition 3 has been satisfied. (Examples have been chosen from the optimization using the first version of the model, i.e. p=0).

Iterate	Data Set 1 ($M)	Data Set 2 ($M)
0	3.676	5.361
20	3.040	4.787
40	2.970	4.635
60	2.928	4.535
80	2.902	4.486
100	2.896	4.424
120	2.859	4.410
140	2.841	4.381
160	2.828	4.372
180	2.820	4.356
200	2.789	4.340
220	2.782	4.319
240	2.779	4.309
260	2.774	4.302
280	2.768	4.299
300	2.763	4.296
320	2.758	4.292
340	2.754	4.288
360	2.750	4.285
380	2.746	4.284
400	2.742	4.281
420	2.739	4.279
440	2.736	4.278
460	2.731	4.277
480	2.725	4.277
500	2.725	4.276

Table 1 Iterates from the System optimization of the Adelaide Network. (Both sets of data/models).

OD Pair k	Chain Flows (Erl) h_1^k	h_2^k	h_3^k	Marginal Chain Costs D_1^k	D_2^k	D_3^k
1	5.56	0.00	5.23	722.9	735.5	724.6
2	3.21	0.00	1.61	420.4	428.1	421.0
4	0.00	0.00	0.85	*****	2508.0	2467.8
527	3.09	1.70	0.00	1196.9	1196.9	1285.6
556	0.00	14.70	0.00	685.7	645.9	746.7
636	1.13	0.14	0.95	2565.6	2565.6	2565.6
697	0.00	0.49	0.00	*****	3048.3	3230.5
1135	0.00	0.00	0.79	4162.5	3751.1	3631.1

***** These OD pairs do not have direct routes.

Table 2 Selected OD pair chain flows and Marginal chain costs for System optimized solution.

3.2 USER OPTIMIZATION

The computer program for determining the User optimizing chain flow pattern for the Adelaide network was started at the point "all flow on the viable direct routes - flow on second choice routes otherwise". The reason for choosing this point is related to Result 5 from Section 2.3. A brief survey of the cost figures for the Adelaide network revealed that this result applied to approximately 200 OD pairs, and thus by choosing this point the chain flows for these OD pairs would be set and would not alter during the computations. This meant that the program could concentrate on those OD pairs which did not satisfy this result and a User solution would be found more rapidly than if it had started from some arbitrary flow pattern.

A solution satisfying the User equilibrium conditions (Definition 4) to within a reasonable tolerance was obtained after 1¼ minutes of computer time on the CDC Cyber 74, and 250 iterations. Table 3 gives the network cost for each of the two optimizations at each twentieth iterate. Table 4 lists selected OD pair chain flows and their corresponding chain costs per unit flow.

Iterate	Data Set 1 ($M)	Data Set 2 ($M)
0	3.676	5.361
20	3.466	5.242
40	3.399	5.240
60	3.360	5.220
80	3.323	5.217
100	3.278	5.215
120	3.257	5.214
140	3.246	5.214
180	3.238	5.214
200	3.228	5.214
220	3.228	-
240	3.220	-
260	3.219	-

Table 3 Iterates from the User optimizations of the Adelaide Network (both models).

OD Pair k	Chain h_1^k	Flows h_2^k	(Erl) h_3^k	Chains costs per unit flow $\bar{C}_1^k$	$\bar{C}_2^k$	$\bar{C}_3^k$
1	10.79	0.00	0.00	426.74	446.72	450.41
2	4.82	0.00	0.00	259.33	353.18	310.58
4	0.00	0.85	0.00	******	1578.48	1919.68
72	1.48	0.11	0.74	364.80	363.27	363.32
85	0.00	0.00	0.44	******	1475.29	1467.95
201	2.73	0.00	0.71	681.41	721.65	681.37
259	0.00	0.06	0.03	******	4170.55	4170.75
1135	0.56	0.23	0.00	2931.5	2931.46	3062.78

Table 4 Selected chain flows and Marginal chain costs for the User optimization of the Adelaide Network. (Data Set 1).

****** These OD Pairs do not have direct routes.

4. COMPARISON OF SYSTEM AND USER NETWORKS

It is interesting to note that the User and System optimized networks exhibit a number of important features. In both studies the unrounded costs of \$M3.22 and \$M5.21 for the User optimized solutions were further away from the unrounded costs of \$M2.73 and \$M4.28 of the System optimized solutions than had been anticipated from results with smaller networks (see Ref. 1). The rapidity with which the User solution was obtained, demonstrates the possible potential which this type of optimization has for iterating towards a System optimized solution.

An important comparison between the User and System solutions concerns the relative numbers of unused links required in each type of network. For the first set of data, the User optimized solution had 35 links on the alternate routes which were unused, while for the System optimized network there were only 20 unused links. Furthermore, there are only 7 of these links which have been rejected by both optimal solutions. One very obvious feature which is common to both solutions is that they have almost invariably been rejecting the first link of a second choice chain, i.e, link 2 in Figure 1.

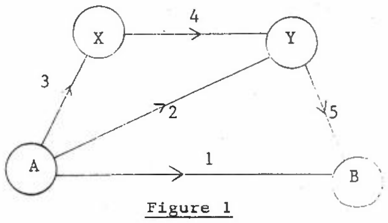


Figure 1

Chain	Route (A-B)
1	1
2	2, 5
3	3, 4, 5

Table 5 Chains for Figure 1

This feature appears to be very significant and a number of suggestions have been put forward to explain this result. Firstly, an examination of the relative costs of the second and third choice chains tends to indicate that the third choice routes have lower costs per circuit than expected in practice. Secondly, third choice chains tend to be attractive for flow because they provide traffic with the opportunity to combine with traffic from as many as ninety origin-destination pairs. This effect has tended to draw traffic away from the second choice routes also, and hence the overall effect on the network is to provide first and third choice routes and limit the provision of links on second choice chains to OD pairs which have relatively expensive first choice routes.

Table 6 compares the costs of direct and alternate routes in each of the optimal network designs. It will be clear from this table that the System optimized networks make more efficient use of the alternate routes, while the User solutions make more extensive use of the direct routes, i.e. the User solutions fail to recognise the advantages of bulking traffic on the alternate routes.

A question of considerable importance concerns the uniqueness of the User optimal solutions obtained in this study. It will be evident from Result 5 in Section 2 of this paper that a network which satisfies condition (17) for all OD pairs will have a simple unique solution. However, in practical networks such as the Adelaide Network, only 209 of the 1141 OD pairs in Data Set 1 satisfy this Result, and this raises the question of uniqueness for such a network. It is necessary to consider the uniqueness of the link flow pattern in the network, not the chain flow pattern, as it is well known that in general, many different chain flow patterns can give rise to the same link flow pattern. Preliminary investigations into the uniqueness of the User solution for the Adelaide Network have tended to suggest that there is only one link flow pattern which produces this solution. For example, when the User optimization process is started from a link flow pattern which has a lower cost than the User solution, it is found that User optimization applied to this pattern results in a return to a link flow pattern which resembles that obtained in the original User optimization. Investigations which have been carried out on this network have so far failed to find any other User optimized link flow patterns, although it is intended to pursue this point further in due course.

Data Set 1 Costs (Rounded)	System Optimal Solution (\$)	User Optimal Solution (\$)
Direct Routes	1,755,054	2,504,742
Alternate Routes	991,342	714,523
Data Set 2 Costs Rounded	**System Optimal Solution (\$)**	**User Optimal Solution (\$)**
Direct Routes	1,909,805	3,705,576
Alternate Routes	2,464,680	1,534,814

Table 6 Comparison of System and User Optimal Solution Costs for Direct and Alternate Routes.

5. CONCLUSIONS

In this paper two concepts of optimality have been applied to Berry's model for dimensioning alternate routing networks. Experience with small networks had suggested that the User optimized network cost differed from the System optimal network cost by one or two percent. The results of applying these concepts to a large realistic network, presented in the foregoing discussion, show that the difference in network costs is greatly increased, and the design of the User network is radically different from the System optimized design.

Results presented in the earlier sections show that a User optimized solution to the Adelaide Network was achieved more rapidly than a corresponding System optimized solution. (This is largely due to the fact that the costs per unit flow are simpler to evaluate than the marginal chain costs). Although the User optimization is more rapid, it is now clear that the speed of this method is not an advantage in obtaining a "near optimal" System solution because of the large differences in cost and design of the User optimized network. As a result of this observation, other methods of rapid System optimization have been tried, and the algorithm of Reference 8 has proved to be quite successful.

Although the ultimate aim of network planners has been to produce a minimum cost alternate routing network, there may be some advantages in considering network designs similar to those produced from User optimization. For example, System optimized networks are very sensitive to overloads since circuits are operating at high efficiencies, however, a User optimized network is not as sensitive to overload and circuits are not operating at such high efficiencies in the alternate routes. Furthermore, the User solution tends to relieve the strain of high traffic loads on the tandem exchanges, and this may be considered desirable in exchanges which are nearing their designed capacities.

6. ACKNOWLEDGEMENTS

The author wishes to express his appreciation for the guidance and encouragement given by Professor R.B. Potts and Dr. L.T.M. Berry of the Department of Applied Mathematics, University of Adelaide. Thanks are also due to Dr. C.W. Pratt and Mr. K.D. Vawser of Telecom Australia for many useful discussions.

7. REFERENCES

1. Harris, R.J., "Concepts of Optimality in Alternate Routing Networks", 7th ITC, Stockholm, June 1973, reprinted in A.T.R. Vol. 7, No. 2, 1973.

2. Harris, R.J., "The Modified Reduced Gradient Method for Optimally Dimensioning Telephone Networks", in A.T.R. Vol. 10, No. 1, 1976.

3. Wolfe, P., "Methods of Non-Linear Programming", in Recent Advances in Mathematical Programming (Graves and Wolfe - editors), pp.60.

4. Berry, L.T.M. and Harris, R.J., "A Simulation Study of the Accuracy of a Telephone Network Dimensioning Model", in A.T.R. Vol. 9, No. 2, 1975.

5. Berry, L.T.M., "An Explicit Formula for Dimensioning Links Offered Overflow Traffic", in A.T.R. Vol. 8, No. 1, pp.13-17, 1974.

6. Berry, L.T.M., "An Application of Mathematical Programming to Alternate Routing", A.T.R. Vol. 4, No. 2, 1970, pp.20-27.

7. Rapp, Y., "Planning of a Junction Network in a Multi-Exchange Area. I. General Principles", Ericsson Technics, Vol. 20, No. 1, 1964, pp.77-130.

8. Harris, R.J., "Concepts of Optimality in Alternate Routing Networks", Thesis for Ph.D., Adelaide University, 1974.

Discussion

C.W.A. JESSOP, Australia :

(i) *The paper states that the cost of a final choice is less than expected in practice. Please explain how this circumstance eventuates.*

(ii) *What method do you propose for manual adjustment of the optimised network should this be necessary due to changes in traffic levels or network configurations. In particular ensuring that OD grades of service are maintained.*

R.J. HARRIS, Australia :

(i) *In the first data set it has been found that some relay set costs have been omitted in some instances and also, allowance was not always made for the heavier gauge cable required for final choice routes.*

(ii) (a) *For changes in traffic levels I would suggest that the chain flows which are affected could be increased (or decreased) in proportion to the changes in these levels. Iterations could then be recommenced from this new "near optional" point and terminated when a suitable accuracy had been obtained for the new minimum.*

(b) *Several different types of changes in network structure are possible. Firstly, if the routing strategies for particular OD pairs are to be altered I would suggest that the old chain flows could be used as a starting solution in the new network arrangement (where possible).*

Once again, iterations would need to be performed in order to re-optimize the network. Secondly a new exchange (or group of exchanges) could be added or deleted. In such a case, the extent to which this change affects the network would be a prime factor in determining the best way to proceed. If changes were extensive it may be more practical to restart the entire optimizing process from the beginning. For minor alterations it should generally be possible to retain old values of the chain flows and simply re-assign these values to the new routes and recommence iterations until a sufficiently accurate solution has been obtained.

C.W.A. JESSOP, Australia : *In Section 4 Para. 5 you state that User optimised solutions fail to recognise the advantages of bulking traffic on alternate routes. However this method rejects 1st Alternate routes in favour of bulking traffic on the final routes. Could you explain this apparent contradiction please.*

R.J. HARRIS, Australia : *Both solutions recognise the advantages of bulking the traffic in the alternate routes and the differences are largely differences in degree, i.e. In the System optimised network, extensive use is made of the alternate routing section of the network; while in the User solution the majority of the traffic remains on the first choice routes and the remaining traffic flows onto the route with the most attractive cost per unit erlang - which would appear to be the third choice chain in these particular instances.*

MUN CHIN, Australia : *How does the Chain Flow method compare with the traditional method (i.e. Pratt et al) in terms of computing time.*

R.J. HARRIS, Australia : *(Adelaide Network) At each iteration step of the System optimization process it is necessary to dimension the network from the current chain flow pattern, this process takes 0.2 seconds CPU time on a Cyber 74 computer. For the same computer, one dimensioning step in the conventional approach can take from 15 to 30 seconds to perform depending on the methods used to determine variances and equivalent random traffic parameters.*

M. ANDERBERG, Sweden : *You state in Section 3.1 of your paper that the iterative process took approximately 15 minutes of CP time, and then you arrived at a solution which was 1-2% from the true System optional solution.*

- *How is this true solution defined.*
- *How long will it take to find the true solution for the studied network.*

R.J. HARRIS, Australia :

(i) *A "true" System optimized solution satisfies equations (5) of my paper given as Definition 3.*

(ii) *In general, a non linear programming algorithm will not necessarily converge to the exact minimum in a finite number of iterations although at each iteration the new chain flow pattern will define a network of lower cost. It is possible to compute a lower bound on the value of the minimum cost from duality theory of non linear programming. In practice, it is necessary to terminate computations once the difference between the current cost and the estimate for the minimum is less than a prescribed figure.*

W. LÖRCHER, Germany : In your paper you compare two optimizing principles : System and user optimization. You compare these two principles by means of a real network and you obtain differences of about 15% with respect to the costs.
My question is:
Have you compared these results also with results obtained by conventional methods.

R.J. HARRIS, Australia : I have compared my results with the conventional methods using Data set 2 but the comparisons should be treated with caution for two reasons:

(i) The two models have different grade of service criteria and it is difficult to assess the OD grades of service from the conventional model.

(ii) The conventional methods minimize OD costs in a "competitive" manner which is related to the concept of Game Theoretic Optimization discussed in Ref. 1.

The results of this comparison show that the cost of a conventional network falls about midway between the System and User solutions. A number of interesting observations can be made concerning the comparison between the System and the conventionally optimized networks. In particular, the numbers of direct route circuits were roughly comparable in the 2 networks, and hence the direct route costs are comparable but the cost of the alternate routing networks differed by about 8-10 percent and circuits were assigned to links in different ways. In some particular cases for Berry's model third choice routes had more circuits than the conventional approach - for reasons cited in Section 4.

A.H. FREEMAN, Australia : In Australian metropolitan networks the use of routes X-B has generally been avoided because it makes optimisation using cost factors unstable. As your results show that A-Y routes are often not provided in this particular network there may be savings in allowing X-B routes. What do you expect in terms of increased computer time and possible non uniqueness of solutions if this routing is permitted in your optimisation procedure.

R.J. HARRIS, Australia :

(i) It is possible that X-B routes may be advantageous in this situation. The input data requires that the network routing pattern must be specified for the optimizing model thus we would need to rerun the System optimizing program to establish the new solution and determine whether such routes would be more attractive to flow than A-Y routes.

(ii) If we retain the requirement of 3 choices for traffic from a given OD pair computing times would not vary greatly from those which were obtained incorporating A-Y routes. If we permit all 4 possible routing strategies the computing times would probably increase by about 20 percent per iteration of the System optimizing algorithm.

(iii) One of the difficulties of a non linear programming formulation is that optimizing algorithms may not necessarily locate the global optimal solution if the objective function is non convex. As some of the link cost functions in the dimensioning model are not convex with respect to the chain flows it is not clear that the total network cost will necessarily be convex. Studies of the total network cost for networks with A-Y routing permitted have not indicated non convexity in the objective function when Data set 1 was employed. We have not done any studies on the use of X-B routing and so we cannot comment upon the possible existence of multiple solutions due to non convexity of the objective function.

J. HARRINGTON, Australia : I refer to Section 5 of the paper in which it is suggested that the network would stand up to overload conditions better under "User" optimization than it would under "System" optimization. I do not disagree with this but I suggest a more economic solution would be to build up the final routes, links 3, 4 and 5 of fig. 1, of the "System" optimized solution. This might result in unequal but improved overall grade of service for most traffic parcels. This would not be undesirable because I suggest there is no real merit in designing for equal end to end grade of service for all traffic parcels as the customer is more likely to assess the service provided by the telephone network on the networks ability to handle, successfully, the first repeat attempt call after congestion was encountered, than he would be on knowing the network was designed for overall grades of service of 1 in or 100 or 1 in 200, etc.
Could you comment please.

R.J. HARRIS, Australia : I am afraid that I cannot agree with your statement concerning the "lack of merit" in using end to end grade of service for at least two reasons.

(i) The conventional approach of specifying grades of service on backbone routes ensures that all OD pairs get service of better than (or equal to) that grade of service. In fact, if the majority of the traffic parcels are carried on the early choice routes, then there is almost no loss for such parcels. This must result in more circuits being provided overall than a method which dimensions for OD grades of service and thus I suggest there are economic advantages in applying the principle of end to end grade of service.

(ii) I also suggest that subscribers are unable to distinguish between grades of service of 0.01 or 0.005 (say) under normal conditions. Furthermore, Berry's model gives the planner the opportunity to specify end to end grades of service for particular OD pairs, thus it is possible to design a network where these OD pairs have different grades of service to meet particular requirements.

It is not clear from your comments how you propose to "build up" the backbone routes from the System optimised solution. Presumably this must be done in some regular and ordered way which avoids specifying circuit quantities which does not lead to large imbalances in OD grades of service - such as the present conventional scheme.

There are many possibilities which could be considered in order to reduce the sensitivity of the System optimized network to overload. One suggestion is to reformulate the constraints of the model in such a way that overload is considered explicitly: e.g. A certain percentage increase in offered traffic load should not result in degradation in OD grade of service of more than a specified value, and then find a network optimal solution for this condition. An alternative suggestion involves the provision of service protection routes for those OD pairs which would normally be offered to the backbone routes only.

J.S. HARRINGTON, Australia :

(i) Are the costs given in table 1 for integer value of circuits?

(ii) Do the unrounded costs in Section 4 paragraph 1 refer to costs for non-integer circuit values?

(iii) What is the reason for the small cost difference, for data set 2, between the values given in Section 4, paragraph 1 and those given in table 6.

R.J. HARRIS, Australia :

(i) The costs given in Table 1 refer to unrounded circuit values.

(ii) Yes, the unrounded costs in Section 4 para. 1 do refer to costs for non-integer circuit values.

(iii) A typographical error in paragraph 1 is responsible for the cost difference. Please note that the figures given in Table 6 are the correct figures.

J.S. HARRINGTON, Australia : *Refer please to Section 4 paragraph 2. I would like to make a comment on the rejection of the second choice route - link 2 of fig. 1, in preference to the third choice route - links 3 and 4 of fig. 1. The probability of this occurring increases:*

(i) *as the cost of the two paths tend towards one another,*

(ii) *if the own area tandem is an XY tandem as against an 'X' and a 'Y'. (i.e.) one inlet as against two plus a R/S.*

(iii) *As the efficiency of the final route increases (more offered traffic and full availability switching).*

(iv) *As the number of terminals served by a 'Y' tandem decreases, i.e. more 'Y' tandems.*

R.J. HARRIS, Australia : *In general I would agree with the various possibilities which you suggest, in fact all of them appear to be evident in the network which I have considered. Note however that (i) occurs as a consequence of possibility (ii) in your list of suggestions.*

BIOGRAPHY

R.J. HARRIS graduated from Adelaide University with a B.Sc. (Hons.) degree in 1971 and commenced work for the degree of Doctor of Philosophy early in 1972 under the supervision of Professor R.B. Potts and Dr. L.T.M. Berry. His field of research involved the study of optimization techniques and their applications to alternative routing networks. In June 1973 he presented a paper at the Seventh International Teletraffic Congress which was based on his post-graduate research. In January 1974 he completed his thesis and joined the Australian Post Office as a Research Officer in the Traffic Engineering Section of Central Office, where he is continuing his research into telephone network optimization, dimensioning techniques and other traffic engineering problems.

A Method for Determining Optimal Integer Numbers of Circuits in a Telephone Network

L. T. M. BERRY
University of Adelaide, Adelaide, Australia

ABSTRACT

For a given telephone network, there exist many different junction allocations which achieve specified overall traffic congestions between each pair of exchanges. This paper considers the problem of finding a Minimum Cost network, that is, a network which satisfies the performance criterion at a minimum total junction cost. Previous models have relaxed the integrality condition on junction numbers.

1. INTRODUCTION

The minimum cost telephone network problem has been formulated as a mathematical program [1,2]

$$\text{Minimize } C(\underset{\sim}{h})$$
$$\sum_{j=1}^{j(k)} h_j^k = b^k t^k \; ; \quad k = 1,\ldots,K. \tag{1}$$
$$h_j^k \geq 0.$$

The variable h_j^k is the mean carried traffic on the jth route (or chain) between the kth origin-destination pair of exchanges. There are $j(k)$ distinct permissible routes for OD pair k. A feasible point $\underset{\sim}{h}$, with non-negative elements h_j^k, represents a chain flow pattern on the network with the property that for each OD pair k the sum of the carried traffic is a prescribed fraction b^k of the total offered traffic t^k. The planned OD traffic congestion g^k $(=1-b^k)$ may be chosen to give different performance priorities to the OD pairs.

For a full availability network, a mathematical model has been developed [1,3] which gives the number of circuits $\underset{\sim}{n}$ required on the links of the network as a function of the chain flow pattern $\underset{\sim}{h}$. $C(\underset{\sim}{h})$ is the total circuit cost. The nonlinear program (1) has been solved using cost and traffic dispersion data for the Adelaide metropolitan network. Two successful solution techniques have been reported [4,5]. Both approaches assume $\underset{\sim}{n}$ is a continuous function of $\underset{\sim}{h}$ and require some kind of rounding to integer numbers at the completion of the optimisation.

Whereas previous optimisation methods have relaxed the integrality condition on the number of circuits, this paper describes a method for determining optimal integer numbers of circuits.

2. PROBLEMS ASSOCIATED WITH THE DIRECT INTEGER FORMULATION

As an integer program, with variables n_i denoting the number of circuits on link i, the minimum cost problem may be formulated

$$\text{Minimize } \underset{\sim}{c}'\underset{\sim}{n}$$
$$\sum_{j=1}^{j(k)} h_j^k(\underset{\sim}{n}) = b^k t^k \tag{2}$$
$$\underset{\sim}{n} \geq \underset{\sim}{0}$$
$$n_i \text{ integer.}$$

It is usual to consider constant costs per unit circuit, c_i. Thus the objective function representing total circuit cost, unlike that in (1), is a linear function. On the other hand the performance constraints lose the simple form of (1) and are nonlinear functions of n_i.

The techniques used to solve (1) were successful because of the echelon-diagonal structure of the linear chain flow constraints. Not only are the corresponding constraints in (2) nonlinear, there is the added difficulty of having a non-convex set of feasible $\underset{\sim}{n}$. This undesirable property is illustrated by the following example. Consider a network with a single OD pair, having two routes to the destination D. Let the components of $\underset{\sim}{n}$ denote the number of circuits on links 1 and 2 respectively. If the point $(n_1,0)$ gives the required OD congestion $E_{n_1}(t^1)$*, by symmetry $(0,n_1)$ is also a feasible point. But as

$$E_{n_1}(t^1) \neq 2E_{n_1/2}(t^1/2)$$

the point $\underset{\sim}{n} = (n_1/2,\, n_1/2)$ is <u>not</u> a feasible point.

Let us next consider the functions $h_j^k(n)$. Assuming random traffic t^k is offered to n_k junctions on the direct route for OD pair k, the functions h_1^k are readily obtained from the Erlang loss formula

$$h_1^k(n_k) = [1 - E_{n_k}(t^k)]t^k. \tag{3}$$

The non-randomly distributed overflow calls offered to second choice routes of the network compete for access to common links and produce the chain flows $h_2^k(\underset{\sim}{n})$. But, although it is possible to estimate the total carried traffics on the common link (for example by application of the equivalent random method) this total cannot be apportioned to the individual streams with sufficient accuracy. Similarly no sufficiently accurate model exists at the present time to allow $h_3^k(\underset{\sim}{n})$ to be determined as a function of the junction vector $\underset{\sim}{n}$.

It is concluded then, that a direct formulation in terms of numbers of junctions leads to problems of considerable difficulty. No general purpose integer programming algorithm exists at the present time for solving large linear programs. We have a large nonlinear program for which the constraint set is non-convex. In addition, there is the further problem that although $\underset{\sim}{n}$ may be determined as a function of $\underset{\sim}{h}$ the inverse function is not known.

3. AN EXTERIOR PERIODIC PENALTY FUNCTION TECHNIQUE

It was observed, when solving problem 1, that rounding the junction numbers to integer values at intervals of 20 successive iterates gave a monotonic decreasing sequence of total cost values for the first 100 iterations. This suggests that the nonlinear function $C(\underset{\sim}{h})$ is 'well behaved' and that a reasonable approximation to the solution to problem (2) may be found by solving (1) and rounding-off circuit numbers to the nearest integer. It may be thought that a safer approach would be to round-up n_i to the next integer.

To examine the sensitivity of OD congestions to changes of 1 circuit on the last links of final choice routes a number of simulation tests were made with a 12 OD pair alternative routing network. It was found that increases of 1 circuit on selected links could result in an in-

* Erlang loss formula

crease in traffic congestion of 20% or more for certain OD pairs. Thus rounding-up $\underset{\sim}{n}$ (obtained as a function of $\underset{\sim}{h}^*$, the solution to (1)) breaks the performance constraints. We consider an alternative approach.

Problem (2) is equivalent to

$$\text{Minimize } C(\underset{\sim}{h})$$
$$\sum_{j=1}^{j(k)} h_j^k = b^k t^k \; ; \; k = 1,\ldots,K. \qquad (4)$$
$$h_j^k \geq 0$$
$$n_i \equiv (\underset{\sim}{n}(\underset{\sim}{h}))_i \quad \text{integer.}$$

To maintain feasibility we must consider only those chain flow patterns $\underset{\sim}{h}$ for which n_i are integers. The author recalls the difficulty of finding, by tedious adjustments, a set of flows $\underset{\sim}{h}$ during a simulation experiment in which the calculated numbers of circuits were required to be integers (to within 2 decimal places).

Instead, the integrality constraints are incorporated implicitly in a penalty function $\phi(n_i)$ which adds a positive value to the objective function whenever some n_i is not an integer. That is, we replace (4) by

$$\text{Minimize } C(\underset{\sim}{h}) + \sum_{i=1}^{I} p_i \, \phi(n_i(\underset{\sim}{h})) \qquad (5)$$
$$\sum_{j=1}^{j(k)} h_j^k = b^k t^k \; ; \; k = 1,\ldots,K. \qquad (6)$$
$$h_j^k \geq 0, \qquad (7)$$

I being the total number of links in the network. The parameters p_i are simply non-negative weights.

Clearly, the function $\phi(n_i)$ must be periodic, with period 1, and it is desirable for ϕ to achieve a maximum value whenever the fractional part of n_i is 0.5. In this paper results are reported for the case

$$\phi(n_i) = \sin^2 \pi n_i \,. \qquad (8)$$

Let S be the set of points satisfying (6) and (7), and define

$$F(p) = \min_{\underset{\sim}{h} \in S} . [C(\underset{\sim}{h}) + \sum_{i=1}^{I} p_i \, \phi(n_i(\underset{\sim}{h}))]. \qquad (9)$$

Suppose, for some sequence of $\underset{\sim}{p}$ values with each component p_i tending to ∞, that $F(\underset{\sim}{p})$ approaches a finite value L and that this value is achieved by the point $\underset{\sim}{h}^*$. It is clear from (9) that each n_i must be integral otherwise $\phi(n_i(\underset{\sim}{h}^*))$ is positive and $F(\underset{\sim}{p})$ could be made arbitrarily large. It should be noted that in general L is not unique, that is, two different sequences of $\underset{\sim}{p}$ values may be chosen giving different limiting values.

Before outlining the proposed solution procedure in full detail the following simple example is given to illustrate certain difficulties.

$$\text{Minimize } \pi(x-1.5)^2 \qquad (10)$$
$$x \text{ integer.}$$

Equation (9) gives for each $\underset{\sim}{p}$ the unconstrained minimization problem

$$F(p) = \min . \pi(x-1.5)^2 + p \sin^2 \pi x. \qquad (11)$$

A necessary condition for a stationary point is that the derivative with respect to x of the objective function for the unconstrained problem be zero. That is,

$$3 - 2x = p \sin 2\pi x. \qquad (12)$$

It can be seen from fig.1 that the solution to (12) is <u>not</u> unique, its value depending on both the starting point for the minimization and the sequence of values for p.

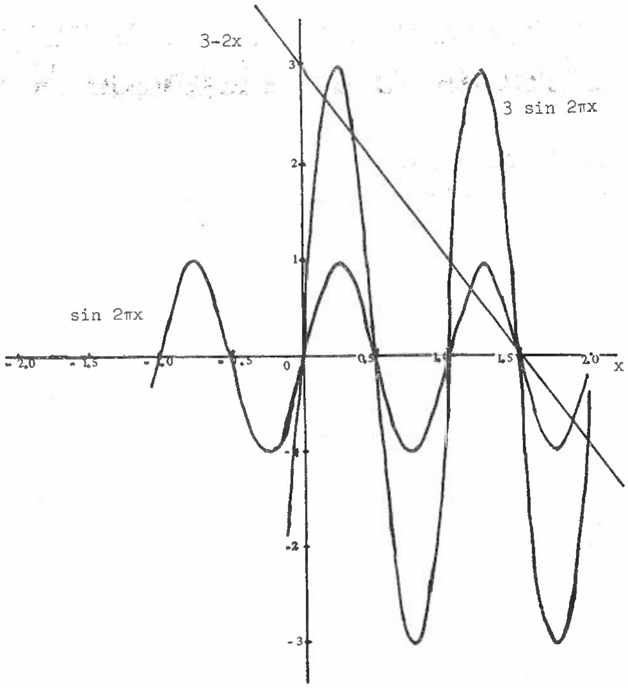


Fig.1 Graphical solutions to equation (12).

For the case p=1 there are two points satisfying (11). If we choose our starting point for the minimization less than 1.5 we are likely to obtain the solution $x_1^* = 1.422$ (4 sig.figs.). Solving (11) for increasing values of p we approach the value x*=1, a solution of (10). Table 1 shows the convergence with a sequence of increasing values of p.

p	1	3	4	10	100
i	1	2	3	4	5
x_i^*	1.422	1.289	1.231	1.015	1.002

Table 1 A sequence of values of p giving convergence to a solution of (10).

It is clear that as p increases the number of solutions to (12) increases; many of these solutions correspond to local minima of (11). For example, with starting point x = 15.4 and p = 10,000 Powell's algorithm finds the solution x*=7, a local minima of (11). To solve the above problem it is necessary to start either with a small value of p or an initial value of x near the point x = 1.5 (the solution to the continuous problem, i.e. p=0).

We now consider problem (2). The proposed approach is to find a near-optimal solution to the program given by (5), (6) and (7) with $p_i = 0$ for all i. This point is then taken as our initial point. The parameters p_i are increased to some small positive value and a new point determined by applying the minimization algorithm (in our case the Gradient Projection algorithm) for a selected number of iterations. p_i are subsequently further increased and the procedure repeated until the solution vector $\underset{\sim}{h}$ leads to values of n_i sufficiently close to integers.

It is reasonable to expect, in many applications, that the objective function is roughly quadratic near its minimum. The technique described provides a means of moving to an integer point near the point which solves the continuous problem. In other words, for our problem, the technique provides an automatic means of adjusting flows on links of the network to obtain integral numbers of circuits. We are able to use the same minimization algorithm which solves (1), only minor changes in the objective function and gradient subroutines being required.

It is interesting to contrast the technique with the sequential unconstrainted minimization techniques proposed by Fiacco and McCormick [6]. The approach is an exterior penalty function method, that is, feasibility (n_i integer) is not maintained. A sequence of points is found converging to a feasible point. The auxiliary function, however, is not monotonic. We have a periodic auxiliary function. Also, whilst it is usual to obtain a sequence of minimizing points, it has been found that it is not necessary to solve (9) to find minima for each $\underset{\sim}{p}$.

4. RESULTS

The telephone network considered has 1141 OD pairs with 755 direct junctions and 212 overflow junctions. The mathematical program has 3,037 variables h_j^k, 1141 performance constraints and 967 integrality constraints.

With very large non-linear programs it is not usually possible to achieve <u>the</u> optimal solution. Some convenient stopping criterion must be determined. Let

$$d_i = \min(f_i, 1-f_i) \tag{13}$$

where f_i is the fractional part of n_i. The algorithm is terminated when the following inequality is satisfied

$$S(\underset{\sim}{d}) \equiv \sum_{i=1}^{I} d_i < K. \tag{14}$$

The value of $S(\underset{\sim}{d})$ gives a convenient measure of the 'distance' from an integer point $\underset{\sim}{n}$.

If the number of circuits in the network is large we may expect, for a continuous variable optimisation, that the values d_i are approximately uniformly distributed between 0 and 0.5 with mean 0.25. Hence $S(\underset{\sim}{d})$ would have an expected value 0.25I. In our case this value is 241.75. The constant K=7 was arbitrarily chosen in the optimisation.

The objective function has a maximum value greater than $4,000,000 (corresponding approximately to all flow on overflow routes). It is known from previous computation, for the continuous problem, that the minimum value is approximately $2,800,00. The starting point for application of the solution technique outlined above gave a total circuit cost of $3,325,469. The parameters p_i were set to the same value, PY, for all direct circuits and also a common value PZ for all overflow circuits. After 300 iterations with PY=15, PZ=10, a point was obtained for which $S(\underset{\sim}{d})$ was 138 (compare 241.75) corresponding to a circuit cost of $2,872,685. This indicates simultaneous movement towards an integer point $\underset{\sim}{n}$ and a decrease in cost. This last point was stored and a number of different strategies considered for choosing successive values of PY and PZ. The most rapid convergence was obtained with the following approach. PY was increased to 1000, PZ remaining 10 and a further 10 iterations performed. This forced n_i close to integer values for direct circuits. Minor adjustments were then made to the flows on direct routes to achieve integer numbers of circuits according to the Erlang loss formula. These flows (h_1^k) were maintained in subsequent iterations. PZ was next increased to 1000 and after a further 49 iterations the stopping criterion was satisfied, $S(\underset{\sim}{d})$ being 6.98.

The final point had a cost of $2,967,533 made up of a cost of $1,677,093 for direct circuits and $1,290,440 for overflow circuits. Theoretically the assigned OD congestions are satisfied, and the mean deviation of the values n_i from integer values is 0.007. The savings achieved ($406,069) represent a decrease of approximately 12% from the starting point. A comparison of the values, at the 300th and 359th iteration, of n_i for the first 100 overflow circuits is given in Tables 2a and 2b.

4.8	37.5	34.2	30.5	21.2	36.8	71.0	51.6	43.7	38.1
45.0	32.8	8.1	61.0	49.5	64.0	63.8	13.7	27.9	54.1
66.7	61.3	13.7	55.1	32.3	74.7	32.2	64.4	82.7	57.6
15.2	28.2	4.2	21.7	35.5	34.6	32.5	34.4	75.3	14.9
20.6	22.0	7.9	16.0	2.0	62.2	16.8	6.6	17.3	5.9
2.4	30.2	0.0	5.4	0.0	2.3	20.6	0.9	0.0	0.0
10.7	12.3	8.1	8.0	5.7	0.6	27.9	17.7	13.5	0.3
0.5	17.5	3.8	10.2	0.0	3.2	30.1	0.0	3.6	0.0
2.0	13.7	2.9	5.5	1.0	9.5	12.1	40.3	6.5	9.2
15.3	8.6	44.4	7.9	27.9	21.5	3.2	19.8	13.7	12.7

Table 2a Values for overflow link numbers n_i (i=1,..,100) at iteration 300.

6.0	39.0	29.0	31.0	17.0	36.0	68.0	51.0	43.0	39.0
40.0	27.0	8.0	39.0	45.0	68.0	58.0	14.0	26.0	53.0
59.0	46.0	12.0	39.0	27.0	73.0	28.0	46.0	80.0	43.0
14.0	24.0	2.1	22.0	33.1	26.9	29.9	34.9	65.0	15.0
21.0	20.0	8.9	8.0	3.0	43.0	17.0	7.0	9.0	6.1
4.0	16.0	1.0	5.0	0.1	1.0	21.0	1.0	0.0	0.0
5.0	7.0	8.0	5.0	5.0	1.0	28.0	18.0	14.0	0.1
1.0	16.0	4.0	8.0	0.1	1.0	30.0	1.0	4.0	0.1
2.0	13.0	3.0	6.0	0.0	3.0	10.0	38.0	7.0	10.0
15.0	9.1	44.1	7.0	20.0	20.0	4.0	21.0	15.1	13.0

Table 2b Values for overflow link numbers n_i (i=1,..,100) at iteration 359.

5. CONCLUSIONS

A technique has been described for determining the optimal integer numbers of circuits in an alternative routing network, subject to the constraint that each pair of OD exchanges in the network experience a prescribed traffic congestion. The problem formulated as a nonlinear program has the structure

$$\begin{gathered} \text{Min. } f(\underset{\sim}{x}) \\ g(H^{-1}(\underset{\sim}{x})) \quad (\overline{\leqslant}) \; \underset{\sim}{b} \\ x_i \quad \text{integer,} \end{gathered} \tag{15}$$

where $H^{-1}(\underset{\sim}{x})$ is not known, but $\underset{\sim}{x} = H(\underset{\sim}{y})$ is known.

The problem was reformulated as

$$\begin{gathered} \text{Min. } f(H(\underset{\sim}{y})) + \sum_{i=1}^{I} p_i \, \phi((H(\underset{\sim}{y}))_i) \\ g(\underset{\sim}{y}) \; (\overline{\leqslant}) \; \underset{\sim}{b}, \end{gathered} \tag{16}$$

where $\phi(t)$ is an appropriate periodic penalty function, and the parameters p_i are increased during the minimization. Although the choice of both a sequence of p_i and the number of iterates between changes in their values is somewhat of an art, it has been demonstrated using real data that the method converges satisfactorily.

An extension of the method to restrict the number of circuits for each link to multiples 0,L,2L,... may be accomplished by choosing a suitable function ϕ with period L, for example $\sin^2 \pi n/L$.

6. BIBLIOGRAPHY

[1] Berry, L.T.M., "An Application of Mathematical Programming to Alternate Routing", Australian Telecommunication Research, 1970, 4, [2], 20-27.

[2] Berry, L.T.M., "A Simulation Investigation of an Alternate Routing Model," Australian Telecommunication Research, 1971, 5, [2], 3-11.

[3] Berry, L.T.M. "An Explicit Formula for Dimensioning Links Offered Overflow Traffic," Australian Telecommunication Research, 1974, 8, [1], 13-17.

[4] Berry, L.T.M., "The Gradient Projection Vector for Multicommodity Network Problems," Australian Telecommunication Research, 1975, 9, [1], 18-27.

[5] Harris, R.J., "The Modified Reduced Gradient Method for Optimally Dimensioning Telephone Networks". To appear in Australian Telecommunication Research, 1976, 10, [2].

[6] Fiacco, A.V.; McCormick, G.P., "Nonlinear Programming: Sequential Unconstrained Minimization Techniques," Wiley, 1968.

Discussion

R.B. LEIGH, U.K. : *Quite rightly the paper takes into account monetary as well as circuit savings. My question concerns this aspect and is in three parts:*

(i) *Why in your presentation did you assume the overflow junctions to be half the price of direct circuits.*

(ii) *In Section 4 of the paper what account has been taken of circuit group modularity in assessing circuit costs, and*

(iii) *Where does the tandem switching cost enter the equation.*

L.T.M. BERRY, Australia : *(i) In the example each overflow chain cost was \$102 per circuit. The direct links had a cost of \$100 per circuit. These costs were chosen arbitrarily (with overflow route costs slightly greater than direct route costs) only to illustrate that network cost is a function of the chain flow pattern. (ii) The actual cost data used in the study of the metropolitan Adelaide network was provided by the Adelaide traffic engineering section of Telecom Australia. A constant average cost per circuit on each link was assumed the last paragraph of the paper indicates how circuit numbers may be restricted to modules in the optimization. A constant cost per module could be assumed in the formulation. (iii) The tandem switching costs were included in the circuit costs. (Mr. Leigh was introduced to Mr. K. Vawser for finer detail).*

G.D. BOLAM, Australia : *In your analysis you seem to ignore the cost of the tandem exchanges, for example you quoted \$100 for a direct circuit and \$51 for an alternate circuit. The cost of an exchange building (and land) could contribute more than \$1000 per alternate circuit : taking this into account would seriously alter your conclusions. The cost of switching equipment is additional. The cost of providing an additional tandem exchange in Melbourne would be quite prohibitive at the present time so an upper limit on alternative routes should appear in your calculations. Please comment.*

L.T.M. BERRY, Australia : *Please refer to the answer given to Mr. R.B. Leigh's question. The above costs were arbitrarily chosen simply to illustrate that network cost is a function of the chain flow pattern. The actual costs used when the model was applied to the Adelaide telephone network were realistic costs - provided by the Traffic Engineering section of Telecom Australia. These costs included switching costs at tandems.*

The model takes a fixed network topology, i.e. the routing pattern and tandem exchanges are fixed. We do not consider the provision of additional tandems.

R. KRISHNAN IYER, Australia : *In your paper you have briefly described an approach to integer optimization. Another approach to integer (discrete) optimization is Dakin's Tree Search Algorithm (modified by Bandler and Chen) which begins by finding a continuous solution. If the solution is non integer, partitioning is introduced and corresponding constraints added. The process continues until an integer (discrete) solution within some limits is obtained. We have worked with this approach and have found that there is good convergence. Is it possible to apply such an approach to your model.*

L.T.M. BERRY, Australia : *No, for the following reasons. To apply either a tree search or branch and bound algorithm it would first be necessary to formulate the problem completely in terms of the numbers of circuits n_i. However no satisfactory model is known which gives the OD traffic congestions as a function of the vector of circuit numbers $\underline{n}$. Apart from this, there are the added difficulties that these constraints are non-linear and define a non-convex region of feasible solutions. The non-linear programme is also very large.*

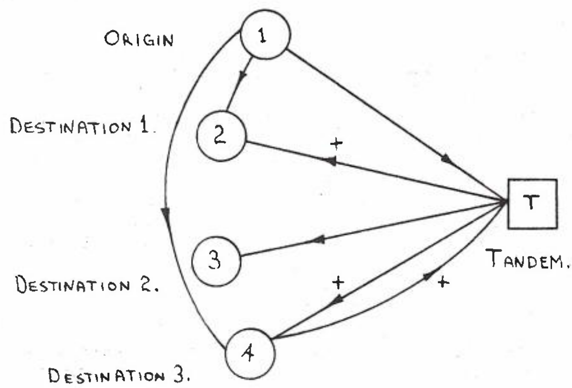
The formulation considered has both continuous variables h_j^k related linearly to OD congestions and discrete variables n_i. When the constraints involving n_i are incorporated in the objective function it is possible to solve very large minimum cost problems by either of the methods given in refs. (4) and (5).

J.P. FARR, Australia : *In the results presented in your paper have you calculated the variance of traffics offered to the alternative routes before computing the circuits required on these routes? If not, the traffic congestion will not be identical for each origin - destination pair as you have claimed.*

L.T.M. BERRY, Australia : *The dimensioning model described in references (1) and (3) uses both means and variances of traffic in the network. The accuracy of the model is discussed in reference (2) and also in reference (4) of paper 512. The statistical analysis of the simulation tests supports the assertion that specified OD congestions can be achieved.*

W. LÖRCHER, Germany : *In your paper you give an example that an increase of 1 circuit on a link could result in an increase in probability of loss of 20% or more for certain OD-pairs. Please can you explain this effect.*

L.T.M. BERRY, Australia :



Traffic between (1)-(3) small (no direct link provided). The addition of 1 circuit on link T - (2) results in an increase in the carried traffic on link (1) - T destined for exchange (2) with a resultant decrease in the carried traffic on chain (1) - T - (3). This effect (viz. the increase in traffic congestion for traffic from exchange (1) to exchange (3) is further compounded by the addition of an extra circuit on link T - (4). Similarly an increase on link (4) - T increases the chain flow between exchanges (4) and (3) but reduces the chain flow between exchanges (1) and (3).

BIOGRAPHY

LES BERRY was educated at the University of Adelaide and Adelaide Teachers' College. For 10 years from 1959-68 he taught in secondary schools. He then joined the South Australian Institute of Technology as a tutor with the aim of working towards a higher degree. In 1972 he was awarded his Ph.D. for a thesis entitled "A Mathematical Model for Optimizing Telephone Networks". Subsequently he joined the Department of Applied Mathematics at the University of Adelaide as a lecturer.

A Probabilistic Model for Optimisation of Telephone Networks

R. S. KRISHNAN IYER
T. DOWNS
University of Queensland, Brisbane, Australia

ABSTRACT

In this paper a probabilistic model for a telephone traffic network is developed. A system of state equations is presented for the basic building block of an alternate routing network and an explicit solution is given. From this model, the blocking probabilities at any network node may be readily obtained. As a consequence, it is possible to formulate an optimization problem for the minimization of the variance of the traffic arriving at the X - tandem (subject to cost and any other required constraints).

NOTATION

D_{ij}	j^{th} destination in the Y_i tandem area.
O_{kl}	l^{th} origin parented on the k^{th} X-tandem X_k.
a_{ij}	Proportion of calls from O_{kl} that are destined for D_{ij}.
m_{ij}	Number of circuits busy on the direct route from the origin O_{kl} to D_{ij}.
$\overline{m}$	Vector containing all m_{ij} in the form $m_{11}m_{12} \ldots m_{1s_1}, m_{21}m_{22} \ldots m_{2s_2}, \ldots, m_{y_1}m_{y_2} \ldots m_{ys_y}$.
m_{Y_i}	Number of busy circuits on the route from O_{kl} to the Y_i tandem.
$\overline{m}_Y$	Vector containing all m_{Y_i}.
$m_{X_{kl}}$	Number of busy circuits on the $O_{kl} \rightarrow X_k$ route.
$m_{X_kY_i}$	Number of busy circuits on the $X_k \rightarrow Y_i$ route.
$m_{Y_{ij}}$	Number of busy circuits on the $Y_i \rightarrow D_{ij}$ route.
$\beta_{ij}(m_{ij})$	Probability of blocking on the direct route to D_{ij} given m_{ij} circuits are busy.
$\beta_{Y_i}(m_{Y_i})$	Probability of blocking on the $O_{kl} \rightarrow Y_i$ route given m_{Y_i} circuits are busy.
$\beta_{X_{kl}}(m_{X_{kl}})$	Probability of blocking on the $O_{kl} \rightarrow X_k$ route given $m_{X_{kl}}$ circuits are busy.
$\beta_{X_kY_i}(m_{X_kY_i})$	Probability of blocking on the $X_k \rightarrow Y_i$ route given $m_{X_kY_i}$ circuits are busy.
$\beta_{Y_{ij}}(m_{Y_{ij}})$	Probability of blocking on the $Y_i \rightarrow D_{ij}$ route given $m_{Y_{ij}}$ circuits are busy.
$P(\overline{m},\overline{m}_Y,m_{X_{kl}})$	Probability that the network is in state $\overline{m}$, $\overline{m}_Y$, $m_{X_{kl}}$.
$P(\overline{m}_{ij} - 1)$	Probability $P(\overline{m},\overline{m}_Y,m_{X_{kl}})$ with the ij component of m equal to $m_{ij} - 1$.
$P(\overline{m}_{ij} + 1)$	Probability $P(\overline{m},\overline{m}_Y,m_{X_{kl}})$ with the ij component of $\overline{m}$ equal to $m_{ij} + 1$.
$P(\overline{m}_{Y_i} - 1)$	Probability $P(\overline{m},\overline{m}_Y,m_{X_{kl}})$ with the Y_i component of m_Y equal to $m_{Y_i} - 1$.
$P(\overline{m}_{Y_i} + 1)$	Probability $P(\overline{m},\overline{m}_Y,m_{X_{kl}})$ with the Y_i component of m_Y equal to $m_{Y_i} + 1$.

1. INTRODUCTION

A prerequisite for the formulation of any optimization problem is a satisfactory mathematical model. The state equation representation of a traffic system is useful for this purpose if an explicit and unique solution to the state equations can be found. In this paper we present the state equations for the basic alternate routing model shown in fig. 1; an explicit solution to these equations is also given. The approach taken here is based upon the work of Whitaker [1] on multi-server, multi-queue systems.

In fig. 1 the broken lines are intended to indicate the presence of many X and Y tandems and many origins and destinations.

2. THE STATE EQUATIONS

In fig. 1, the node O_{kl} is one of r origins (the l^{th}) parented onto the k^{th} X - tandem X_k. The node D_{ij} is one of s_i destinations (the j^{th}) in the Y_i tandem area. We assume that the network contains a number x of X-tandems and y of Y-tandems.

The traffic originating at O_{kl} destined for D_{ij} takes the direct route to D_{ij} if a line is available; if this route is blocked, it overflows onto the i^{th} Y - tandem route and if this route is also blocked it overflows onto its X - tandem route X_k. If this final route is blocked, the call is lost.

In formulating the mathematical model we assume the following:

(i) The traffic arriving at any origin O_{kl} is Poisson.

(ii) The holding times of calls are distributed exponentially.

(iii) The process is in statistical equilibrium.

(iv) In the infinitesimal time interval dt, no more than one call arrives or is terminated.

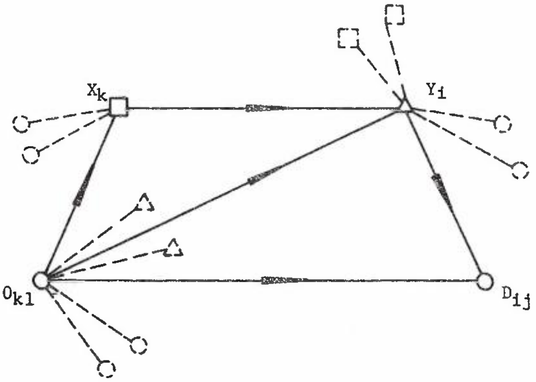


Fig. 1

The state equilibrium equation is written out below and followed by an explanation.

$$
\begin{aligned}
&\left\{\sum_{i=1}^{y}\sum_{j=1}^{s_i}\left(a_{ij}(1-\beta_{ij}(m_{ij}))\right) + \sum_{i=1}^{y}\left((1-\beta_{Y_i}(m_{Y_i}))\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right)\right.\\
&\quad + \left(\sum_{i=1}^{y}\beta_{Y_i}(m_{Y_i})\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right)\left(1-\beta_{X_{k1}}(m_{X_{k1}})\right)\\
&\quad \left. + \sum_{i=1}^{y}\sum_{j=1}^{s_i} m_{ij} + \sum_{i=1}^{y} m_{Y_i} + m_{X_{k1}}\right\} P(\overline{m},\overline{m}_Y,m_{X_{k1}})\\
&= \sum_{i=1}^{y}\sum_{j=1}^{s_i}\left(a_{ij}(1-\beta_{ij}(m_{ij}-1))\right)P(\overline{m}_{ij}-1)\\
&\quad + \sum_{i=1}^{y}(1-\beta_{Y_i}(m_{Y_i}-1))P(\overline{m}_{Y_i}-1)\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\\
&\quad + \left(\sum_{i=1}^{y}\beta_{Y_i}(m_{Y_i})\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right)\left(1-\beta_{X_{k1}}(m_{X_{k1}}-1)\right)P(m_{X_{k1}}-1)\\
&\quad + \sum_{i=1}^{y}\sum_{j=1}^{s_i}(m_{ij}+1)P(\overline{m}_{ij}+1)\\
&\quad + \sum_{i=1}^{y}(m_{Y_i}+1)P(\overline{m}_{Y_i}+1)\\
&\quad + (m_{X_{k1}}+1)P(m_{X_{k1}}+1)
\end{aligned}
\qquad (1)
$$

The left hand side of this equation gives the probability of transition out of the state $\overline{m}$, $\overline{m}_Y$, $m_{X_{k1}}$ and the right hand side gives the probability of transition into this state. Since the terms involved in both sides are similar we give an explanation for one side only.

The first term on the right hand side represents a transition into the state $\overline{m}$, $\overline{m}_Y$, $m_{X_{k1}}$ due to an arrival and no blockage on the direct routes. The second term represents a transition into $\overline{m}$, $\overline{m}_Y$, $m_{X_{k1}}$ due to an arrival and no blockage on the Y_i routes. The third term represents an arrival and no blockage on the X_{k1} route. The remaining three terms represent a transition into state $\overline{m}$, $\overline{m}_Y$, $m_{X_{k1}}$ due to the termination of a call on the direct routes, Y_i routes and X_{k1} route respectively.

3. SOLUTION OF THE STATE EQUATIONS

Equation (1) describes a Birth and Death process where the number of states is finite. The solution of the equation is therefore unique and describes a genuine probability distribution [2]. It is given by

$$
\begin{aligned}
&P(\overline{m},\overline{m}_Y,m_{X_{k1}}) =\\
&= \left\{\frac{1}{m_{X_{k1}}!}\left(\sum_{i=1}^{y}\beta_{Y_i}(m_{Y_i})\sum_{j=1}^{s_i}(\beta_{ij}(m_{ij})a_{ij})\right)^{m_{X_{k1}}}(1-\beta_{X_{k1}}(m_{X_{k1}}-1))(1-\beta_{X_{k1}}(m_{X_{k1}}-2))\right.\\
&\qquad\qquad \ldots(1-\beta_{X_{k1}}(0))\\
&\quad \times \prod_{i=1}^{y}\frac{1}{m_{Y_i}!}\left(\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right)^{m_{Y_i}}(1-\beta_{Y_i}(m_{Y_i}-1))(1-\beta_{Y_i}(m_{Y_i}-2))\\
&\qquad\qquad \ldots(1-\beta_{Y_i}(0))\\
&\quad \left.\times \prod_{j=1}^{s_i}\frac{a_{ij}^{m_{ij}}}{m_{ij}!}(1-\beta_{ij}(m_{ij}-1))(1-\beta_{ij}(m_{ij}-2))\ldots(1-\beta_{ij}(0))\right\}
\end{aligned}
\qquad (2)
$$

We will now show that the steady state solution described by the above relation satisfies the equilibrium equation (2). The following recurrence relations are readily obtained from equation (2):

$$
\begin{aligned}
&m_{X_{k1}}P(m,\overline{m}_Y,m_{X_{k1}})\\
&= \left\{\sum_{i=1}^{y}\beta_{Y_i}(m_{Y_i})\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right\}\left\{1-\beta_{X_{k1}}(m_{X_{k1}}-1)\right\}P(m_{X_{k1}}-1)
\end{aligned}
$$

$$
\begin{aligned}
&m_{Y_i}P(\overline{m},\overline{m}_Y,m_{X_{k1}})\\
&= \left\{(1-\beta_{Y_i}(m_{Y_i}-1))\sum_{j=1}^{s_i}(\beta_{ij}(m_{ij})a_{ij})\right\}P(\overline{m}_{Y_i}-1)
\end{aligned}
$$

$$
m_{ij}P(\overline{m},\overline{m}_Y,m_{X_{k1}}) = \left\{a_{ij}(1-\beta_{ij}(m_{ij}-1)\right\}P(\overline{m}_{ij}-1)
$$

$$
\begin{aligned}
&\left\{\sum_{i=1}^{y}\beta_{Y_i}(m_{Y_i})\sum_{j=1}^{s_i}\beta_{ij}(m_{ij})a_{ij}\right\}\left\{1-\beta_{X_{k1}}(m_{X_{k1}}-1)\right\}P(\overline{m},\overline{m}_Y,m_{X_{k1}})\\
&= (m_{X_{k1}}+1)P(m_{X_{k1}}+1)
\end{aligned}
$$

$$
\left\{(1-\beta_{Y_i}(m_{Y_i}))\sum_{j=1}^{s_i}(\beta_{ij}(m_{ij})a_{ij})\right\}P(\overline{m},\overline{m}_Y,m_{X_{k1}}) = (m_{Y_i}+1)P(\overline{m}_{Y_i}+1)
$$

$$
\left\{a_{ij}(1-\beta_{ij}(m_{ij}))\right\}P(\overline{m},\overline{m}_Y,m_{X_{k1}}) = (m_{ij}+1)P(\overline{m}_{ij}+1) \qquad (3)
$$

By taking the appropriate summations we obtain the desired equilibrium equations.

We have now carried out an analysis of all routes from the origin O_{k1}. The marginal probabilities for these routes may be obtained from the joint probability distribution by summing appropriate terms from the vectors $\overline{m}$ and $\overline{m}_Y$ and the component $m_{X_{k1}}$.

The blocking probabilities B are then determined in terms of their capacities C, from expressions such as the following:

$$B_{ij} = \sum_{m_{ij}=0}^{C_{ij}} \beta_{ij}(m_{ij})P(m_{ij})$$

$$B_{Y_i} = \sum_{m_{Y_i}=0}^{C_{Y_i}} \beta_{Y_i}(m_{Y_i})P(m_{Y_i})$$

$$B_{X_{k1}} = \sum_{m_{X_{k1}}=0}^{C_{X_{k1}}} \beta_{X_{k1}}(m_{X_{k1}})P(m_{X_{k1}})$$

From the marginal probabilities we may also determine the average traffic on each route analysed

$$\text{e.g. } E(m_{ij}) = \sum_{m_{ij}=0}^{C_{ij}} m_{ij}P(m_{ij})$$

$$E(m_{Y_i}) = \sum_{m_{Y_i}=0}^{C_{Y_i}} m_{Y_i}P(m_{Y_i})$$

$$E(m_{X_{k1}}) = \sum_{m_{X_{k1}}=0}^{C_{mX_{k1}}} m_{X_{k1}}P(m_{X_{k1}})$$

4. DETERMINATION OF SIGNIFICANT TRAFFIC CHARACTERISTICS

Now to complete our model we need the probability expressions for the $X_k \to Y_i$ route and the $Y_i \to D_{ij}$ route.

Consider first the $X_k \to Y_i$ route. The proportion of the traffic from O_{k1} to the Y_i tandem area that overflows onto the $O_{k1} \to X_k$ route is given by

$$\alpha(ki)1 = \sum_{j=1}^{s_i} a_{ij}B_{ij}B_{Y_i}(1 - B_{X_{kl}})$$

Thus the average arrival at the X_k tandem of the traffic destined to the Y_i tandem is equal to

$$A_{X_kY_i} = \alpha(ki)1 + \alpha(ki)2 + \ldots + \alpha(ki)r$$

where r is the number of origins.

Once again assuming statistical equilibrium, no time correlation, and that probability of more than one arrival in time dt may be ignored, we may write the probability $P(m_{X_kY_i})$ of the number of calls on the $X_k \to Y_i$ route as follows

$$P(m_{X_kY_i}) = \frac{A_{X_kY_i}^{m_{X_kY_i}}}{m_{X_kY_i}!} \prod_{m_t=0}^{m_{X_kY_i}-1} \left\{1 - \beta_{X_kY_i}(m_t)\right\} P(0)$$

where P(0) is determined by the normalizing condition

$$\sum_{m_{X_kY_i}} P(m_{X_kY_i}) = 1$$

Analysing the $Y_i \to D_{ij}$ route we obtain the following.

The proportion of traffic from $O_{kl} \to D_{ij}$ that reaches Y_i from X_k is given by

$$a_{ij}B_{ij}B_{Y_i}(1 - B_{X_{kl}})(1 - B_{X_kY_i})$$

The proportion of the traffic reaching Y_i from the $O_{kl} \to Y_i$ route is given by

$$a_{ij}B_{ij}(1 - B_{Y_i})$$

.'. the traffic arriving at Y_i from O_{kl} and destined for D_{ij} is given by

$$a_{Y_{kj}} = a_{ij}B_{ij}\left\{B_{Y_i}(1 - B_{X_{kl}})(1 - B_{X_kY_i}) + (1 - B_{Y_i})\right\}$$

Thus we have an expression for traffic reaching Y_i from all X tandems and the total arriving at Y_i and destined for D_{ij} is given by

$$a_{Y_{ij}} = \sum_{k=1}^{X} a_{Y_{kj}}$$

The probability $P(m_{Y_{ij}})$ of the system being in state $m_{Y_{ij}}$ is then

$$P(m_{Y_{ij}}) = \left\{\frac{a_{Y_{ij}}^{m_{Y_{ij}}}}{m_{Y_{ij}}!} \prod_{t=0}^{m_{Y_{ij}}-1} 1 - \beta(t)\right\} P(0)$$

subject to the normalising condition

$$\sum_{m_{Y_{ij}}} P(m_{Y_{ij}}) = 1$$

The blocking probability and average traffic for these routes may be found in exactly the same fashion as those in equations (4) and (5).

For example, the average blocking in the route between Y_i and D_{ij} is given by

$$\sum_{m_{Y_{ij}}} \beta_{Y_{ij}}(m_{Y_{ij}})P(m_{Y_{ij}})$$

and the average traffic on this route is given by

$$\sum_{m_{Y_{ij}}} m_{Y_{ij}}P(m_{Y_{ij}})$$

5. SYSTEM OPTIMIZATION

Now that we have all the significant probabilistic information for the system, we are in a position to formulate an optimization problem. For high efficiency, one would require maximum utilization of the final route. Consequently a reasonable approach would appear to involve minimization of the variance of traffic arriving at each X-tandem. This approach is simplified through the fact that a unique set of origins is parented to each X-tandem. This allows us to choose the sum of the variances of the traffic arriving at all the X-tandems as the objective function in the optimization problem.

The primary constraint on the problem is cost. A second constraint that we impose is on the grade of service at each destination; we require the blocking probability of calls to a destination D_{ij} in the network to be some predetermined value.

The programming of this approach is now under consideration and it is hoped that some preliminary results will be available at the congress.

CONCLUDING REMARKS

In this paper, we have presented a general-purpose model for a metropolitan telephone network. This model should prove useful as a design aid and with appropriate extensions should be valuable in the allocation of direct routes, switching availabilities and service protection routes.

ACKNOWLEDGEMENTS

We should like to thank J. Rubas who introduced us to this problem area and also R.J. Harris, R. Stevens and D.N.P. Murthy for useful discussions.

REFERENCES

[1] B. Whitaker, 'An analysis and optimisation of multiserver, multiqueue systems with finite waiting space in each queue', Ph.D. Dissertation, New York University, 1974.

[2] R. Syski, 'Introduction to congestion theory in telephone systems.' Oliver and Boyd, 1960.

Discussion

M. ANDERBERG, Sweden : *Mr Iyer, in the presentation of your paper, you included some comments on the accuracy of the Rapp's approximation,*

$$A \simeq V + 3\frac{V}{M}\left(\frac{V}{M} - 1\right).$$

(I should perhaps point out that this is the only approximation formula, the value of n is found exactly, when knowing A.) Since it is an approximation and is as such very useful for a starting point in the iterative process of finding A and n, I do not understand your comments on rt's inaccuracy.

R. KRISHNAN IYER, Australia : *In answer to your question, we would like to make the following points.*

We are quite aware of the procedures under which Rapp's approximation is usually used. If equivalent random theory (ERT) is used in the normal fashion, Rapp's approximation is used to determine approximate values of A and n. These values are then adjusted to the correct values by means of an error minimization procedure. The correct values of A and n can then be used in a relation of the form

$$M_{n+1}(A) = \frac{A.M_n(A)}{n+1+M_n(A)}$$

The above form is used C times, where C is the number of circuits from which one wishes to determine the overflow.

This method is well-established, is widely used, and is well-known to give satisfactory results. It is not this procedure which we seek to question, although, before proceeding to our main argument we would like to draw attention to one point of interest.

In the presentation of the paper, we showed that according to the ERT the variance of the overflow traffic is extremely sensitive to the number of trunks, n. Indeed, as was demonstrated in the presentation, a change of less than 0.04% in n can lead to a change of 100% in V. In normal practice this phenomenon does not cause problems, since small changes in n are accompanied by corresponding adjustments in M. However, we feel that it is rather odd that the mathematical model should exhibit such extreme sensitivity, as measured by the first partial derivative.

We would now like to turn our attention to the application of ERT in optimization.

In the optimization process, one is attempting to determine the optimum number of circuits.

As we pointed out in the presentation, we may write

$$V = M\left(1 - M + \frac{A}{1+D}\right)$$

where $D = M+n-A$.

Since $M+n+1 \simeq A$ (as can be seen from Rapp's results), and since, in addition, the difference between M and $\left(1 + \frac{A}{1+D}\right)$ is very small, V is evaluated by forming the difference between two numbers which are close together. Thus V is very sensitive to D.

If this result is applied in a continous optimization procedure, continuous values are obtained for the numbers of circuits. If these values are rounded off, the values of variance, as calculated by ERT, are likely to deviate quite considerably from the optimum on each link.

For example, assume one is trying to optimize two link variables X and Y. Suppose, for the purposes of our argument, that the optimum numbers of trunks are 25.6 and 26.7. If these values are to be rounded off, one cannot be sure how the variance is going to vary from its value at the optimum.

It has been suggested that an integer optimization procedure be employed. It is not clear to us how such a procedure will overcome the sensitivity problem. For instance, if the integer solution obtained for the above example were 24, 25, then, once again, there is no indication of how the variance has deviated from the optimum.

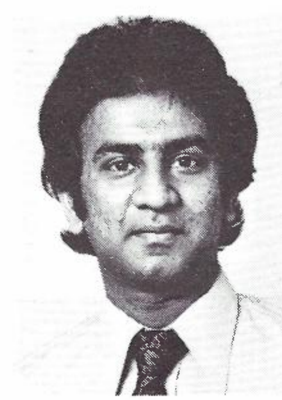
Since ERT is a method which takes the second moment into account, we feel that procedures which use ERT and which do not take proper account of the second moment require further investigation.

M. ANDERBERG, Sweden : In your paper you introduce a system of state equations. Considering the amount of states that will occur, it would be of great interest to hear of the preliminary results, envisaged in your paper, of the programming of your formulae and the use on some networks.

THOMAS DOWNS, Australia : Unfortunately, our answer is not a very satisfactory one since we do not have any preliminary results. However, there are two major reasons why this is so. Firstly, as Mr. Krishnan Iyer mentioned, we recently found that our model could be modified in order to eliminate special assumptions on traffic arrivals at tandem exchanges. Secondly, our interest has been recently directed toward the problem of formulation of switching strategies, which, in the simple form outlined by Mr. Krishnan Iyer, we feel may constitute a more attractive starting point for the implementation of our model. We agree that there is a large number of network states but for the formutation of switching strategies it seees likely that only a small number of state probabilities need be considered. In addition, a close examination of the rather forbidding looking explicit solution reveals that a large amount of computation is common to a large number of states and the savings in computational effort which can be achieved by taking advantage of this fact remain to be seen.

BIOGRAPHY

THOMAS DOWNS was born in Huddersfield, England on 27 June, 1946. He received the degrees B. Tech. and Ph.D. from the University of Bradford, England in 1968 and 1972, respectively. From 1968 until 1973 he was employed by the Marconi Company in Chelmsford, England in the Theoretical Sciences Laboratory where he was involved in the computer-aided analysis of electrical networks and systems. In 1973 he was appointed Lecturer in Electrical Engineering at the University of Queensland. His current interests are operations research in Electrical Engineering, numerical problems in circuit and system theory and the computation of electromagnetic fields. Dr. Downs is a member of IEEE, associate member of IEE and associate Fellow of the Institute of Mathematics and its Applications.



BIOGRAPHY

R. KRISHNAN IYER was born in New Delhi, India on 4 December, 1949. He received the B.E. degree from the University of Queensland, Australia in 1972. He is currently working towards a Ph.D. in Electrical Engineering at the University of Queensland. He is studying operations research, especially applications of probability theory and optimization techniques.

AUSTRALIAN
TELECOMMUNICATION
RESEARCH

VOLUME 11, NUMBER 1, 1977

Special Issue — EIGHTH INTERNATIONAL TELETRAFFIC CONGRESS

Contents*

*Abbreviated Titles