THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLV NOVEMBER 1966 NUMBER 9

Programming and Control Problems Arising from Optimal Routing in Telephone Networks*

By V. E. BENEŠ

(Manuscript received June 10, 1966)

In many circumstances a telephone call can be completed through a connecting network in several ways. Hence, there naturally arise problems of optimal routing, that is, of making the choices of routes so as to achieve extrema of one or more measures of system performance, such as the loss (probability of blocking) or the carried load.

As is customary in traffic theory, a Markov process is used to describe network operation with complete information. The controlled system is described by linear differential equations with the control functions (expressing the routing method being used) among the coefficients. Restricting attention to asymptotic behavior leads to a problem of maximizing a bilinear form subject to a linear equality constraint whose matrix is itself constrained to lie in a given convex set. An alternative approach first shows that minimizing the loss, and maximizing the fraction of events that are successful attempts to place a call, are equivalent. This fact permits a dynamic programming formulation, which, in turn, leads to a very large linear programming problem. Two small examples are treated numerically by this method.

It is particularly important to try to verbalize, and then mechanize, the optimal routing strategies. In this endeavor, the linear programming formulation is of limited usefulness. Therefore, in the latter half of the work we

* Presented at the First International Conference on Programming and Control, USAF Academy, Colorado, April 15–16, 1965. A prolonged abstract of the present text appears as part of the proceedings of the Conference in the SIAM Journal on Control, vol. 4, number 1, February 1966, pp. 6–18.

have attempted to use the special combinatorial structure imposed by the telephonic origins of the problem to shed light on the character of the optimal strategies. In particular, we show that for connecting networks with suitable combinatorial properties, the optimal route choices can be very simply described. Some of the results obtained were suggested by, and verify, conjectures from the practical lore of telephone routing.

The problem of routing calls falls into two parts: Which attempted calls should be accepted in which states? What route should an accepted call use? The first problem is very hard, and only sample numerical answers for small networks are obtained. We solve the second problem analytically for a large class of cases by appeal to combinatorial structure in the network. These cases can be described roughly as those in which the relative merit of states (as far as blocking is concerned) is consistent or continuous; i.e., if a state x is "better" than another y, then the neighbors of x are in the same sense "better" than the corresponding neighbors of y. An abundance of examples indicates that these cases are numerous and so warrant attention. In a network with this kind of combinatorial property, a policy which rejects no unblocked calls and minimizes the number of additional calls that are blocked by completing an attempted call differs from an optimal policy only in that the latter may reject some calls.

I. INTRODUCTION

A telephone connecting network invariably provides many paths on which a particular telephone call can be completed. One of the operational problems faced by the control unit of a telephone system is then to assign to each accepted and completable call a path and, in particular, to choose these assigned paths in the best way. This is the problem of optimal routing of telephone calls. Thus, in the theory of telephone traffic there naturally arise mathematical problems of optimal routing, that is, of making choices of routes in probabilistic models for operating networks so as to achieve extrema of well-defined measures of system performance, such as the probability of blocking (loss).

Unfortunately, it is not unfair to state that the voluminous probabilistic theory of telephone traffic, now some sixty years old, still has rather little to say about how routes for calls should be chosen. We are speaking here of the *mathematical theory* of traffic. Naturally, a wealth of useful information about routing has accumulated over the years from experience in the telephone field; recently it has been buttressed and extended by many simulation studies. This information, nevertheless, still lies largely outside the province of the existing theory of telephone traffic.

It is the aim of this work to formulate, study, and (in part) solve a

general class of optimal routing problems for telephone networks. The formulation of these problems is undertaken insofar as possible within the classical dynamical theory of telephone traffic initiated by A. K. Erlang, that is, in terms of Markov processes based on the assumptions of (*i*) negative exponential distributions for mutually independent holding-times, and (*ii*) randomly originating traffic. To these assumptions is added a description of how attempted calls are accepted and assigned routes.

We conclude this introduction with a brief summary of the entire paper. A complete summary appears later (Section IX) after concepts for formulating the problem have been discussed. As is customary in telephone traffic theory, we use a Markov process to describe the operation of the connecting network under study. The Kolmogorov equations for this process then constitute a set of linear differential equations describing the controlled system; in these the control functions expressing the routing method being used appear among the coefficients. It is natural to restrict attention to asymptotic behavior; this leads to a problem of maximizing a bilinear (or linear fractional) form subject to linear constraints; this problem is equivalent to a linear programming problem. An alternative approach first shows that minimizing the probability of loss, and maximizing the fraction of events that are successful call attempts, are equivalent. This fact permits a classical dynamic programming approach. The remainder of the paper attempts to use this approach to establish relations between combinatorial properties of the network and the policy(ies) optimal for given criteria of performance. In particular, it is shown that for connecting networks having certain "monotone" properties, optimal policies for minimizing loss correspond closely to the heuristic advice, "Prefer those states in which as few calls are blocked as possible".

II. INFORMATION FOR ROUTING DECISIONS

The problem of choosing "good" routes for information flow in a communications network is vastly complicated by the difficult questions surrounding the collection, updating, and relevance of information (about the state of the system) on the basis of which routing decisions are to be made. Thus, one of the items to be chosen in designing a routing scheme is the information on which the routing is to be based. Indeed there is a whole spectrum of possible choices for this information, from no information at all (except what is unwittingly discovered in making call attempts), to full knowledge of the state of the connecting network. Clearly, a practical compromise between total ignorance and a

very expensive, complex scheme based on many data must usually be made.

Our considerations in this work will be limited to the case of perfect information, in which the microscopic state of the connecting network is assumed known and available for making routing decisions. This case is, of course, very far from realistic: few existing or envisaged systems utilize even a small fraction of this possible information for routing. Indeed, much of it is likely to be of very little relevance. Nevertheless, it is important to know what would be good routing if we could implement it and could afford it, so the full information case to be considered here forms at worst a limiting situation for which some theory is available, and a natural starting point for investigation.

III. ACCEPTANCE OR REJECTION OF UNBLOCKED CALLS

In the present discussion of the involved problem of routing calls, one of the difficulties that arises deserves special mention. This difficulty is the problem of deciding whether to accept or reject attempted calls which are not blocked.

At first sight, it might seem that no unblocked call attempt should ever be rejected. The natural argument for this view is that the whole point of a telephone system is to complete calls, and that by rejecting an attempt that could have been completed, the system only lowers its performance. Sensible as this argument sounds, it is unacceptable because it turns out that whether rejection of an unblocked call improves or lowers performance depends on the index of performance, on the distribution of traffic among the sources, on the "community of interest" aspects of the system, etc. If the probability of blocking is used as an index, the "bad" effect of adding a particular call in a given state of the system may be so great and so lasting that it is better to reject the call, and improve the chance of completing many later calls.

To put the matter another way, the problem of routing with full information seems at first to boil down to the question: "Which of the paths available for call c in state x should be used?" This form of the problem overlooks the possibility that perhaps the best thing to do when the state is x and c is attempted is not to complete c at all, but to reject it! In other words, it assumes that, naturally, c will be put up in state x if it is attempted in x and is not blocked. This assumption has always been made in previous applications of the model we use.[1,2]

Conceivably, then, it is better to reject a call c that is not blocked in a state x. Thus the problem of routing should be phrased: "Should a call c, free and not blocked in state x, be completed, and if so, by which route?"

It turns out that answering the first part of the question, as to which calls should be completed in which states, is often the hardest part of the problem. Examples can be given in which it is fairly easy to solve the route selection part of the problem, but for which the question of whether a call should go in or not is not settled. That this question has substantial practical import is apparent from the simulation studies carried out by J. H. Weber,[3] which clearly show how in trunking networks prohibition of circuitous routes (and thus rejection of certain unblocked calls) can improve system performance.

J. H. Weber[4] has also remarked that the problem of deciding whether an unblocked call should be refused is closely related to the distinction between *trunking networks*, used in toll systems to interconnect towns and cities, and *central office networks*, used to interconnect trunks and customers' lines at a single location. An important combinatorial difference between the two types of networks depends on whether all calls use the same number of links. This is usually the case in central office networks, but rarely true in trunk networks. One result suggested by this distinction would be that a call should always be put up when all calls use the same number of links, but that circuitous routes might be profitably disallowed otherwise.

It appears then that network structure bears on the problem of what calls to accept. However, examples can be given which show that even when there is almost no network structure, other factors such as the distribution of traffic and the "community of interest" can make rejection of some calls part of an optimal policy.

For example, if two lines calling at rates λ_1 , λ_2 , respectively, compete for one trunk, the probability of blocking is

$$\frac{2\lambda_1\lambda_2}{\lambda_1 + \lambda_2 + 2\lambda_1\lambda_2},$$

if no unblocked call is rejected. If the calls of the line calling at rate λ_1 are always rejected, the probability of blocking (with rejected calls included among the blocked) is

$$\frac{\lambda_1\lambda_2 + \lambda_1}{\lambda_1 + \lambda_2 + \lambda_1\lambda_2}.$$

(We have assumed that all calls have unit mean holding-time.) It follows here that if

$$\lambda_2^2 > \lambda_1 + \lambda_2 + \lambda_1\lambda_2$$

then it is better to reject all λ_1 calls than to put them all in! This example, although somewhat unrealistic, illustrates how the distribution of

traffic affects the rejection problem, even in the absence of network structure.

For an example involving the "community of interest", consider two disjoint sets of $(n + 1)$ lines communicating over one trunk, with the quirk that each set has a distinguished line which only attempts calls to the distinguished line in the other set, while the other n lines of one set only attempt calls to the n nondistinguished lines of the other set. Let c be the call consisting of the two distinguished lines talking to each other. If c is always rejected, the probability of blocking is

$$\frac{1 + \lambda n^2(n-1)^2}{n^2 + \lambda n^2(n-1)^2},$$

where we have assumed that lines which call each other do so at rate λ, and holding-times have unit mean. If c is always accepted when it is not blocked, then the probability of blocking is

$$\frac{2\lambda n^2 + \lambda n^2(n-1)^2}{2\lambda n^2 + 1 + n^2 + \lambda n^2(n-1)^2}.$$

From these formulas it follows that it is better to reject c entirely if n is large enough, or if λ is large enough, while if λ is small enough it is better always to accept c.

IV. STATES, EVENTS, AND ASSIGNMENTS

The elements of the mathematical model to be used for our study of routing separate naturally into combinatorial ones and probabilistic. The former arise from the structure of the connecting network and from the ways in which calls can be put up in it; the latter represent assumptions about the random traffic the network is to carry. The combinatorial and structural aspects are discussed in this section; terminology and notation for them are introduced. The probabilistic aspects are considered in a later section.

A *connecting network* ν is a quadruple $\nu = (G,I,\Omega,S)$, where G is a graph depicting network *structure*, I is the set of nodes of G which are *inlets*, Ω is the set of nodes of G that are *outlets*, and S is the set of permitted *states*. Variables x,y,z at the end of the alphabet denote states, while u and v (respectively) denote a typical inlet and a typical outlet. A state x can be thought of as a set of disjoint chains on G, each chain joining I to Ω. Not every such set of chains represents a state: sets with wastefully circuitous chains may be excluded from S. It is possible that $I = \Omega$, that $I \cap \Omega = \theta$ = null set, or that some intermediate condition

obtain, depending on the "community of interest" aspects of the network ν.

The set S of states is *partially ordered* by *inclusion* $\leqq$, where $x \leqq y$ means that state x can be obtained from state y by removing zero or more calls. If x and y satisfy the same *assignment* of inlets to outlets, i.e., are such that all and only those inlets $u \in I$ are connected in x to outlets $v \in \Omega$ which are connected to the same v in y (though possibly by different *routes*), then we say that x and y are *equivalent*, written $x \sim y$.

The set S of states determines another set $\mathcal{E}$ of *events*, either *hangups* (terminations of calls), *successes* (successful call attempts), or *blocked or rejected calls* (unsuccessful call attempts). The occurrence of an event in a state may lead to a new state obtained by adding or removing a call in progress, or it may, if it is a blocked call or one that is rejected, lead to no change of state. Not every event can occur in every state: naturally, only those calls can hang up in a state which are in progress in that state, and only those inlet-outlet pairs can ask for a connection between them in a state that are idle in that state. The notation e is used for a (general) event, h for a hangup, and c for an attempted call. If e can occur in x we write $e \in x$. A call $c \in x$ is *blocked* in a state x if there is no $y \in S$ which covers x in the sense of the partial ordering $\leqq$ and in which c is in progress. For $h \in x$, $x - h$ is the state obtained from x by performing the hangup h.

We denote by A_x the set of states that are immediately above x in the partial ordering $\leqq$, and by B_x the set of those that are immediately below. Thus,

$$A_x = \{\text{states accessible from } x \text{ by adding a call}\}$$

$$B_x = \{\text{states accessible from } x \text{ by a hangup}\}.$$

For an event $e \in x$, the set A_{ex} is to consist of those states $y \neq x$ to which the network might pass upon the occurrence of e in x. Thus, if e is a blocked call, $A_{ex} = \{\theta\}$; also

$$\bigcup_{h \in x} A_{hx} = B_x$$

$$\bigcup_{\substack{c \in x \\ c \text{ not blocked in } x}} A_{cx} = A_x.$$

The *number of calls in progress* in state x is denoted by $|x|$. The number of call attempts $c \in x$ which are not blocked in x is denoted by $s(x)$, for "*successes in x*." The functions $|\cdot|$ and $s(\cdot)$ defined on S play important roles in the stochastic process to be used for studying routing.

It can be seen, further, that the set S of states is not merely partially ordered by $\leqq$, but also forms a semilattice, or a partially ordered system with intersections, with $x \cap y$ defined to be the state consisting of those calls and their respective routes which are common to both x and y. (See G. Birkhoff,[5] p. 18, ex. 1 and footnote 6.)

An *assignment* is a specification of what inlets should be connected to what outlets. The set A of assignments can be represented as the set of all fixed-point-free correspondences from I to Ω. The set A is partially ordered by inclusion, and there is a natural map $\gamma(\cdot)\colon S \to A$ which takes each state $x \in S$ into the assignment it realizes; the map $\gamma(\cdot)$ is a semilattice homomorphism of S into A, since

$$x \geqq y \quad \text{implies} \quad \gamma(x) \geqq \gamma(y),$$

$$\gamma(x \cap y) \leqq \gamma(x) \cap \gamma(y).$$

V. ROUTING MATRICES

It will be assumed throughout this work that attempted calls to busy terminals are rejected, and have no effect on the state of the network; similarly, blocked attempts to call an idle terminal are refused, with no change of state. Attempts to place a call are completed instantly with some choice of route, or are rejected, in accordance with some policy of routing.

Two mathematical descriptions of how routes are assigned to calls will be used. The first, the *routing matrix*, is convenient for writing the Kolmogorov equations for the Markov processes representing network operation. The second, called a *policy*, affords a convenient notation for the actual determination of optimal routing methods for various networks to be described in detail later. Either description is a *rule* or *doctrine* for routing.

A routing matrix $R = (r_{xy}), x,y \in S$, has the following properties: for each $x \in S$, let Π_x be the partition of A_x induced by the equivalence relation $\sim$ of "having the same calls up," or satisfying the same assignment of inlets to outlets; then for each $Y \in \Pi_x$, r_{xy} for $y \in Y$ is a *possibly improper* probability distribution over Y, (that is, it may not sum to unity over Y),

$$r_{xx} = s(x) - \sum_{y \in A_x} r_{xy},$$

and $r_{xy} = 0$ in all other cases.

The interpretation of the routing matrix R is to be this: any $Y \in \Pi_x$ represents all the ways in which a particular call c not blocked in x

(between an inlet idle in x and an outlet idle in x) *could* be completed when the network is in state x; for $y \in Y$, r_{xy} is the chance that if this call c is attempted in x, it will be completed by being routed through the network so as to take the system to state y. That is, we assume that if c is attempted in x, then with probability

$$1 - \sum_{y \in A_{cx}} r_{xy} \tag{1}$$

it is rejected (even though it is not blocked), and with probability r_{xy} it is completed by being assigned the route which would change the state x to y, for $y \in A_{cx}$. The possibly improper distribution of probability $\{r_{xy}, y \in Y\}$ indicates how the calling rate λ due to c is to be spread over the possible ways of putting up the call c, while the improper part (1) is just the chance that it is rejected outright.

This description of routing matrices is a generalization of that used in Refs. 1 and 2 in that it permits, in the nonvanishing of (1), the rejection of unblocked calls forbidden in the cited references.

Thus, a routing matrix R is any function on S^2 with $r_{xy} \geqq 0$, $r_{xy} = 0$ unless $y \in A_x$ or $y = x$, and such that

$$r_{xx} = s(x) - \sum_{y \in A_x} r_{xy}$$

and

$$\sum_{y \in A_{cx}} r_{xy} \leqq 1,$$

for all $c \in x$ not blocked in x. A routing matrix corresponds to a *fixed rule* if $r_{xy} = 0$ or 1 for $x \neq y$; otherwise it corresponds to a *randomized rule*. The convex set of all possible routing matrices is denoted by C.

A *policy* is a function φ: $\mathcal{E} \times S \to S$ such that $c,h \in x$ imply

$$\varphi(c,x) \in A_{cx} \cup \{x\}$$

$$\varphi(h,x) = x - h.$$

It is apparent that a policy is equivalent to a fixed rule; the circumstance that $\varphi(\cdot,x)$ is defined also for hangups h is useful in the sequel. Variables φ,ψ are used to denote policies.

The routing rules and doctrines that might be considered here are of course more numerous by far than those we have introduced above. In particular, time-dependent rules and history-dependent rules are natural generalizations. However, since we will be considering only time-invariant traffic and ergodic Markov processes as representations of operating networks, such generalizations add little of significance.

An important point, however, is that the routing methods here considered are based on a complete knowledge of the state of the system, i.e., we postulate that we are in the case of "perfect information." This postulate is grossly unrealistic for present day electromechanical telephone systems; for an electronic system with a very large and very cheap memory, it becomes realistic: the state of the network can actually be stored and the routing rule in use represented by a giant translator. Such a procedure overcomes the obvious impracticality of determining the state by examination of the actual network, and is actually used in the Bell System's No. 1 ESS (Electronic Switching System).[6]

The routing matrices R used in Refs. 1 and 2 had the property that if a call is not blocked in a state, then it is completed in *some* way; *only* blocked attempts or attempts to busy terminals are rejected. Thus none of these rules for routing resembles the methods that are at present likely to be used in practice. However, since C contains rules that reject certain calls in certain states, even though these calls are not blocked, it turns out that a large class of routing rules which do mirror what might happen in practice is included in C.

Some of the simplest routing rules are not based on any knowledge about the current state of the network. Given a call c that has been attempted, they provide a list of routes to be tried in order; the first route found available is used for the call. The list may include all possible routes for c, or only some of them. It is easy to construct a routing matrix to represent such a rule. Let $r_1, r_2, \cdots, r_n$ be the routes to be tried for a call c. For each state x in which c can occur, let $r_{xy} = 1$ if use of the first r_i that is available in x takes the system from x to y, and let $r_{xy} = 0$ for all other $y \in A_{cx}$. If no route for c that is available in x is among $r_1, \cdots, r_n$, then c is rejected in x even though it may not be blocked, simply because the "sieve" for finding routes is too coarse.

It was assumed in the previous paragraph that no information about the state was used. If it is known, e.g., in which element A of a partition Π of S the state currently is, a similar rule can be represented by a class of lists (of routes to be tried in order), one for each $A \in \Pi$. The same kind of construction then yields the appropriate R. Here the A such that $x_t \in A$ is acting as the "information state."

Thus, many R from C which reject certain calls in certain states describe a rule which closely resembles what is done in practice, e.g., in the translator of the Bell System No. 4A crossbar switching system.

VI. PROBABILISTIC ASSUMPTIONS AND STOCHASTIC PROCESSES

A Markov stochastic process x_t taking values on S is used as a mathematical description of an operating connecting network subject to random

traffic. It is assumed that this operation is in accordance with one of the routing matrices R of Section V. The rest of the process x_t is based on two simple probabilistic assumptions:

(*i*) Holding-times of calls are mutually independent variates, each with the negative exponential distribution of unit mean.

(*ii*) If u is an inlet idle in state x, and $v \neq u$ is any outlet, there is a (conditional) probability

$$\lambda h + o(h), \qquad \lambda > 0$$

that u attempt a call to v in $(t,t+h)$ if $x_t = x$, as $h \to 0$.

The choice of unit mean for the holding-times merely means that the mean holding-time is being used as the unit of time, so that only the traffic parameter λ needs to be specified.

It is convenient to collect these assumptions and the chosen routing matrix R into one transition rate matrix $Q = (q_{xy})$ characteristic of x_t : this matrix is given by

$$q_{xy} = \begin{cases} 1 & \text{if } y \in B_x \\ \lambda r_{xy} & \text{if } y \in A_x \\ -|x| - \lambda[s(x) - r_{xx}] & \text{if } y = x \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In terms of the transition rate matrix Q, it is possible to define an ergodic stationary Markov stochastic process $\{x_t, t \text{ real}\}$ taking values on S. The matrix $P(t)$ of transition probabilities

$$P_{xy}(t) = Pr\{x_t = y \mid x_0 = x\}$$

satisfies the equations of Kolmogorov

$$\frac{d}{dt} P(t) = QP(t) = P(t)Q, \qquad Q(0) = I,$$

and is given formally by the formula

$$P(t) = \exp tQ.$$

Since the zero state (the state with no calls in progress) is accessible from any state in a finite number of steps with positive probability, the process has only one ergodic class, and there exists a unique nonnegative row-vector

$$p = \{p_x, x \in S\}$$

such that as $t \to \infty$

$$P(t) \to \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix},$$

and p satisfies the "statistical equilibrium" or stationarity condition $p'Q = 0$, which can be written out in full in the simple form

$$[|\, x \,| + \lambda s(x) - \lambda r_{xx}]p_x = \sum_{y \in A_x} p_y + \lambda \sum_{y \in B_x} p_y r_{yx}, \qquad x \in S.$$

It is possible that a confusion arises in the mind of the reader as to whether we are talking about central office connecting networks or large trunk networks such as the toll system. For in telephone traffic theory these two areas of application are often described by different models: a "finite-source" model like the present one, in which the conditions of the inlets and outlets form a significant part of the state of the system, is commonly used for the former; an "infinite source" model, with groups of customer's lines reduced to Poisson sources of traffic, is frequently used for the latter. The reason for this difference is that it has simply turned out to be sufficient, in the toll case, to restrict attention to the trunking network as the object of principal interest, and to use the simpler Poisson description of sources.

In principle, of course, the model to be used here serves to describe either area listed above, although in the toll case it naturally demands use of a very large number of states. Thus, in the sequel we make no attempt to distinguish the toll case from the central office case. This viewpoint is justified by the fact that the results to be obtained are robust under passage from finite- to infinite-source models, or they can be reformulated and reproved in the infinite-source context.

VII. FORMULATION OF THE ROUTING PROBLEM

The most common figure of merit used by telephone traffic engineers for evaluating connecting networks is the probability of blocking, the fraction of call attempts that are blocked. It is natural, therefore, to use this quantity as the objective function in our optimization problem of routing. It has been shown[2] for the process x_t to be studied here that if no unblocked call is rejected the probability of blocking (in the mnemonic form $Pr\{bl\}$) is given in terms of the stationary state probability vector p by the formula

$$Pr\{bl\} = \frac{\sum_{x \in S} p_x \beta_x}{\sum_{x \in S} p_x \alpha_x} = \frac{p'\beta}{p'\alpha},$$

where

β_x = number of idle inlet-outlet pairs that are blocked in state x,

α_x = number of idle inlet-outlet pairs in state x.

By the same methods it follows that for a process x_t defined in terms of an $R \in C$ the fraction of attempted calls which are not completed (are "lost"), be it because they were blocked or simply rejected, is given by

$$\frac{p'(\beta + r)}{p'\alpha},$$

where $r = \{r_{xx}, x \in S\}$ is the diagonal of the routing matrix R.

We can now replace the informal problem of minimizing, by suitable routing, the fraction of call attempts that are lost by a precise problem of mathematical programming, as follows: Choose $R \in C$ so as to achieve

$$\min \frac{p'(\beta + r)}{p'\alpha}$$

subject to $p'Q = 0$, $p'1 = 1$, and $p \geqq 0$. (The '1' in '$p'1$' is the vector with all components 1.) Of the constraints, the first is the equilibrium condition on p, the second states that the components of p sum to one, and the third says that p is nonnegative. It is understood, of course, that Q is to be related to R by (2) or, what is the same, by

$$Q = H + \lambda R - \text{diag}\,(|x| + \lambda s(x) + 2\lambda r_{xx}) = Q(R),$$

where $H = (h_{xy})$ is the "hangup matrix" such that $h_{xy} = 1$ or 0 according as $y \in B_x$ or not.

Several authors have formulated routing problems for communications systems. Many of these problems have dealt with systems of the store-and-forward type, in which information is alternately stored at and transmitted from a node in the network without setting up a "continuous path" from source to destination. Such formulations are inapplicable to telephone systems. A possible exception, though, is that[7] of R. Kalaba and M. Juncosa which, for a given amount of traffic between each specified source and destination, and a given network having capacity constraints, attempts to find continuous routes that are best in the sense of maximizing the delivered traffic by solving a linear programming problem.

In its possible application to telephony, this model envisions a given traffic pattern (i.e., a description of who wants to talk to whom) to be satisfied at a particular moment, and tries to find a way of routing as much of this traffic as possible through the network. In our terminology, a traffic pattern is an *assignment* $a(\cdot)$, and satisfying it means finding

an $x \in S$ such that $\gamma(x) = a$. The amount of traffic carried is simply the number $|x|$ of calls in progress. Of course, it is not always possible to satisfy an assignment. Thus, Kalaba's and Juncosa's formulation translates into our setup as follows: Given an assignment $a(\cdot)$ either find $x \in S$ with $\gamma(x) = a$, or else if $a(\cdot)$ is unrealizable, find $x \in S$ which realizes as much of $a(\cdot)$ as possible, i.e., such that $\gamma(x) \leqq a$ and $|x|$ is a maximum. This can be rephrased as follows: If $a(\cdot)$ is given, form the cone

$$K = K(a) = \{a_1 : \quad a_1 \leqq a\},$$

and within $\gamma^{-1}(K)$ pick a state x that is maximal in that $|x| \geqq |y|$ for each $y \in \gamma^{-1}(K)$.

It is to be emphasized that this problem is markedly different from our form of the routing problem. The former is purely combinatorial in character. There is no parameter such as the traffic λ per inlet-outlet pair, so the problem involves no probability, and can have nothing to do with the "grade of service" as customarily employed by telephone engineers. Furthermore, the whole formulation overlooks the fact that in present systems call completions must be made without disturbing calls already in progress.

VIII. PRINCIPLES OF ROUTING

It is important to distinguish *methods* of routing from *principles* of routing. A method of routing is a specific way of accepting or rejecting attempted calls and choosing routes in a particular system, e.g., that implicit in the translator of the Bell System No. 4A crossbar switching system. A *principle* of routing is a kind of general prescription of what constitutes* "good" or "optimal" routing; it is the backbone of many routing methods that might be based on it.

A principle of routing is particularly useful if it has two properties:

(*i*) It is relatively simple and intuitive to state.
(*ii*) There is a substantial class of systems for which it describes the (or part of the) optimal routing method.

In our mathematical setting a method of routing corresponds roughly to a rule $R \in C$. We shall see that the "best" rule $R \in C$ can be obtained by solving a linear programming problem. Now if it should happen that for an interesting class of networks the solutions of these linear programs had some common characteristic, some combinatorial property

* Or, more usually, of what someone's intuition tells him constitutes.

of the sets of states of the networks that served as an alternate description of the linear program solution, then this characteristic or property could be abstracted into a genuine *principle of routing.*

Alternatively, one could formulate as conjectures some intuitive principles of routing, and then try to determine for what classes of networks (if any!) these principles did, in fact, describe the optimum routing methods. This second approach will be followed in the present work; the rest of this section is devoted to a discussion of some *a priori* reasonable candidates for "good" routing rules. All of these candidates are expressions of one and the same idea, namely, that one routing rule is better than another if it avoids more "bad" states, where a "bad" state x is one for which β_x is high. This idea is not just an attractive first approximation to "good" or even optimal routing; it leads at once to conjectures for which our results later in the paper provide strong support in precise ways.

In spite of the lack of general theoretical knowledge about routing, traffic engineers have developed various conjectures and intuitive ideas about what might constitute "good" methods for choosing routes. These conjectures are a natural starting place for any rigorous approach to routing, because the formulation of precise theoretical models in which routing can be studied at once raises the question, "Which of these methods, conjectured to be good, can be proved to be optimal in some theoretical model?" Since many of these methods are relatively simple to describe, and hence to mechanize, established answers to this question would have immediate practical applications. Some of these conjectures will now be discussed.

It is apparent that in a telephone system, putting up a new call can only increase the number of idle pairs that are already blocked. Another way of saying this is that in giving service, i.e., in realizing an attempted call in a connecting network, one is possibly denying service to certain inlets and outlets presently idle, who might attempt a call in the very immediate future. This observation has given rise to a number of routing rules (for systems with blocked attempts refused) of great intuitive appeal, which can be described collectively by the admonition: To decrease (minimize?) the probability of blocking, put in new calls in such a way as to minimize the additional congestion resulting from the new calls.

It is illuminating to discuss particular forms of this advice. One form is this: Route new calls through the most heavily loaded part of the network that will accept them. Another is: Put in a given new call so as to minimize the chance that the next attempt to place a call be blocked.

Or: Avoid blocking states, that is, prefer states in which fewer idle pairs are blocked.

For all the intuitive appeal possessed by these rules, rather little is known about them. Nevertheless, they provide conjectures that will be examined in the precise setting of our theoretical model to yield, we hope, the beginnings of a mathematical theory of optimal routing. Let us see what these rules enjoin in terms of our model. If we put up a call c so as to take the system to a state y, the chance that the next event is a blocked call attempt is

$$\frac{\beta_y}{|y| + \lambda\alpha_y}.$$

Suppose that we just left state x, so that $y \in A_{cx}$. This probability will be smallest if y was chosen according to the "maximum $s(\cdot)$" policy, that is,

$$s(y) = \max_{z \in A_{cx}} s(z),$$

i.e., if we prefer states in which fewer idle pairs are blocked. Thus, in our model the second two forms of the above advice coincide.

Another conjecture arises out of consideration of *gradings* in which calls overflowing certain primary routes are pooled and offered to overflow circuits. Here a natural expectation is that one should always "fill the holes in the multiple," meaning by this that a primary route should be used whenever possible, so that the overflow is left available to as many lines as possible. It will be shown for certain examples that if calls are accepted unless they are blocked, then this rule both describes the optimum routing choices, and is equivalent to the "maximum $s(\cdot)$" policy of the previous paragraph.

IX. SUMMARY AND DISCUSSION

In Sections I to VII the problem of routing calls in a telephone network has been formulated as a mathematical one within Erlang's basic traffic theory. Some routing rules which are intuitively reasonable candidates for "good" or even optimal routing were described in Section VIII.

Since the expansion of $\{p_x, x \in S\}$ such that $p'Q = 0$, $p > 0$, is known,[1,2] it is natural to start in Section X with a consideration of $Pr\{bl\}$ for low traffic: $\lambda \to 0$. We have

$$p_x = p_0 \frac{\lambda^{|x|}}{|x|!} r_x + o(\lambda^{|x|}), \qquad \lambda \to 0,$$

where r_x is the number of strictly ascending (in $\leqq$) paths from $\mathbf{0}$ to x which are permitted by R. If x is a blocking state it contributes a term

$$\frac{p_x\beta_x}{p'\alpha} = p_0 \frac{\lambda^{|x|} r_x \beta_x}{|x|!\, p'\alpha} + o(\lambda^{|x|}), \qquad \lambda \to 0$$

to $Pr\{bl\}$ if no calls are rejected. It follows that for sufficiently low traffic the policy that minimizes r_x is optimal within the policies that reject no calls. In a similar way, it can be shown that *always refusing* a call c cannot be optimal for λ sufficiently small, and that there is never any point in rejecting a call attempt in a state x with

$$|x| < \min\{|y|: \quad y \in S, \beta_y > 0\},$$

for λ small enough.

The *nonlinear* problem of choosing R to minimize $Pr\{bl\}$ is reduced to a *linear programming* problem in Section XI. This reduction substantially facilitates obtaining numerical results, examples of which appear later in this summary.

In an effort to identify optimal routing policies, attention now (Section XII) shifts away from the formal linear programming approach to the underlying Markov process. It is shown that minimizing $Pr\{bl\}$, and maximizing the fraction of events which are successful call attempts, are equivalent; this fact leads to a direct dynamic programming approach, in which

$$\min_{R \in C} \quad Pr\{bl\}$$

and

$$\lim_{n \to \infty} \quad n^{-1} \max E\{\text{number of successful call attempts in } n \text{ events}\}$$

(with the maximum in the second expression over all possible policies for n events) are both achieved by essentially the same stationary policies. The word 'essentially' hides the inherent nonuniqueness of optimal policies due to symmetries in the network and to the possible presence of transient states.

In Section XIII it is shown, following C. Derman, that minimum blocking is achieved by a *fixed* rule.

The mathematical programming problems arising in this new approach are again of the linear programming type, and are similar to those arising in Section XI. Our principal interest, however, does not remain with calculating numerical solutions, but shifts abruptly to the relationships of these solutions to the combinatorial structure of the network. Thus,

the second half of this paper consists less of suitable programming problems than of intuition and combinatorics applied to exhibit (*in parte* or *in toto*) the solutions of these problems and their dependence on and origin in network structure.

The attempt to discover and characterize optimal policies in a wholesale way by appeal to network combinatorics (rather than piecemeal by numerical calculation) begins in Sections XIV and XV with consideration of some simple examples; these lead to the introduction of some "monotone" properties (of connecting networks) which impose the condition that (roughly) the relative merit (as far as blocking is concerned) of states is consistent or continuous, i.e., that if a state x is "better" than another y, then the neighbors of x are in the same sense "better" than the corresponding neighbors of y.

Consideration of these properties is justified by the facts that (*i*) they appear in the examples, and (*ii*) they yield a series of closely knit results (Theorems 7–15) that go far to bear out the heuristic guesses in Section VIII about the nature of good routing. In particular, in a network with one of the monotone properties, a policy which rejects no unblocked calls and minimizes the number of additional calls that are blocked by completing an attempted call differs from an optimal policy only in that the latter may reject some calls. In other words, the "max $s(\cdot)$" policy is optimal to within rejection of calls.

Each monotone property gives rise to a corresponding isotony theorem which gives a *numerical* expression to the relative merits of routes for calls that are implicit in the purely *combinatorial* monotone property. The relevance of these isotony theorems to optimal routing is explained heuristically in Section XVI. The theory culminates, in Section XVIII, with two optimal routing theorems based on the monotone properties. When one of these properties obtains, these results completely answer the question: Which route should be used for an accepted call when there is a choice of routes? Determining the extent to which these combinatorial properties occur in networks of interest appears to be the next major problem in any continuation of the present study.

It is to be stressed that the monotone properties we introduce serve only to identify the route that a call should take *if it is to be accepted;* they do not in any way help to decide which calls should be accepted. Except for the low-traffic results of Section X, and the (obvious and easily proved) fact that in a nonblocking network no call should be rejected, the problem of acceptance or rejection of calls remains an enigma. Some light on it is shed by the numerical results that immediately follow this summary.

The paper concludes in Appendix A with the remark that if the performance index is modified so as to put greater emphasis on "early blocked attempts", i.e., ones occurring soon after the system is started, then no calls should be rejected. The result is proved in detail for this index: the expected number of events until the first blocked attempt. Such a criterion corresponds to trying to avoid the undesirable event, the blocked call, as long as possible.

We turn now to numerical results obtained by solving the linear programming formulation of Section XI for two simple networks. The first is the three-stage Clos network with 2×2 switches depicted in Figs. 1 and 2, and already considered as an illustration of routing in Refs. 1 and 2. The second is a 6-line to 4-trunk concentrator in which each line has access to 2 trunks; it is shown in Figs. 3 and 4. In this second case, the probabilistic model was modified to make $\lambda > 0$ the calling-rate per idle line, rather than that per idle inlet-outlet pair.

In each example, both the minimal probability of blocking, and the probability of blocking under random routing, were calculated for several values of λ by use of the LP90 program. To be more precise, two linear programming problems were solved for each example; the first determined the optimal policy, the second determined the optimal policy among those policies which assigned random routes to accepted calls.

Several important qualitative features of the optimal routing policy were the same in both examples and are described together in the following list:

(*i*) The optimal policy rejected no calls.
(*ii*) The routes assigned by the optimal policy coincided with those that keep $s(\cdot)$ as large as possible.
(*iii*) The optimal policy was the same for all values of the traffic parameter λ examined.
(*iv*) The improvement over random routing brought about by optimal routing decreases as the traffic λ increases.

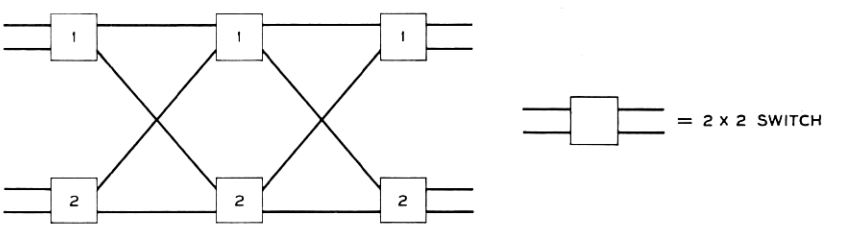

Fig. 1 — 3-stage Clos network with 2×2 switches.

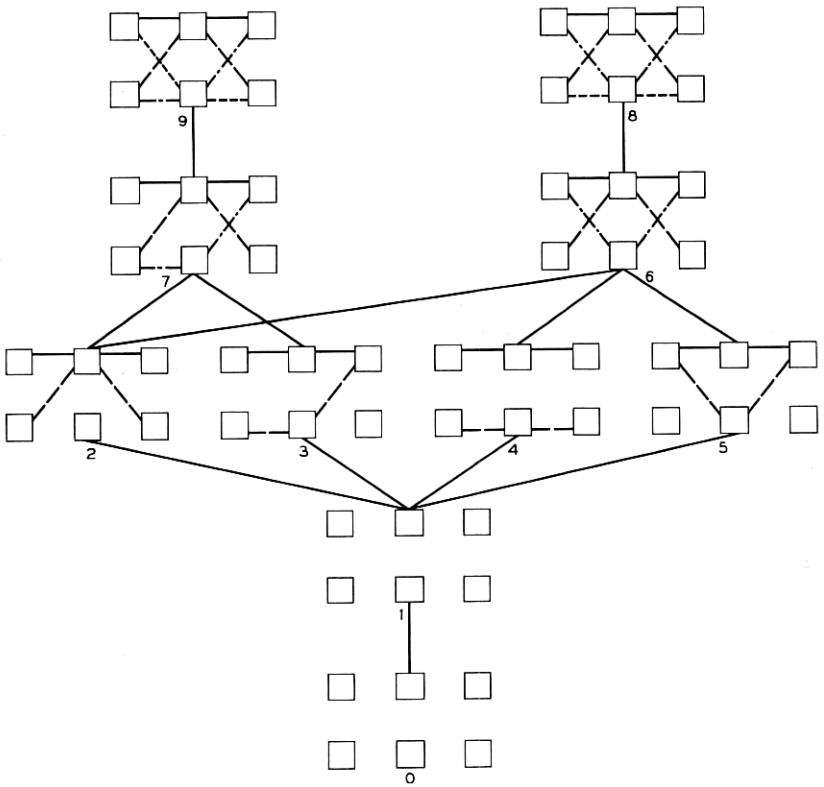

Fig. 2 — States of 3-stage Clos network of Fig. 1.

Under the constraint that accepted calls be routed at random the optimal policy was again to accept all unblocked attempted calls.

Results for the Clos network are given in Fig. 5 and Table I. It is apparent that for low λ optimum routing gives a loss that is easily an *order of magnitude* less than that due to random routing. At high values of λ the difference all but disappears. This behavior is explained in part by the fact that there is no blocking in the "upper" states of Fig. 2; when λ is very large the system spends all its time in these states; when λ is low, however, the occasion for a choice between states 2 and 4 often arises and a correct choice makes a significant difference. (At *very low* values of λ the difference will again decrease because only state 1 will ever be visited with any frequency.)

Results for the concentrator are shown in Fig. 6 and Table II. They include a numerical comparison with hand-calculated loss figures from

unpublished work of S. P. Lloyd dated *circa* 1953. At that time Lloyd studied this particular concentrator model, correctly guessed the optimal policy, proved its optimality for low λ, and calculated the loss for some values of λ. This example exhibits the behavior, conjectured in Ref. 2, p. 275, that a good (here, optimal) policy make certain "bad" states *transient* states. The state numbered 9 is such a transient state under the optimal policy found numerically by the linear programming method.

The present study of routing in telephone networks has suggested a number of conclusions and conjectures:

(*i*) The problem of optimal routing of calls in telephone connecting networks (with full information) can be formulated and solved with Erlang's classical theory of traffic. In this endeavor, the contrasting techniques of machine calculation and combinatorial analysis can be employed either as alternative methods or as complementary approaches.

(*ii*) The problem separates into two parts, that of deciding which calls to accept, and that of choosing routes for accepted calls. Analytically, the first part appears to be much harder than the second, which frequently has a simple intuitive solution closely related to the structure of the network.

(*iii*) Posed within Erlang's theory, the routing problem can be reduced to a (usually very large) linear programming problem and attacked numerically, or studied in terms of Markov decision processes and dynamic programming.

(*iv*) In an apparently wide class of connecting networks, certain natural monotone properties and some isotonies based on them

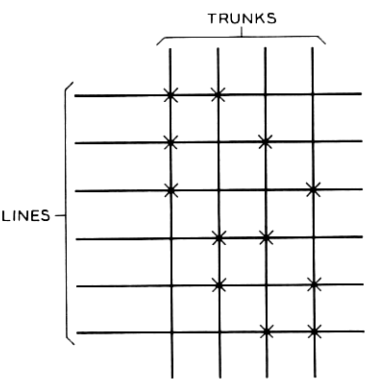

Fig. 3 — 6-to-4, 2 access concentrator.

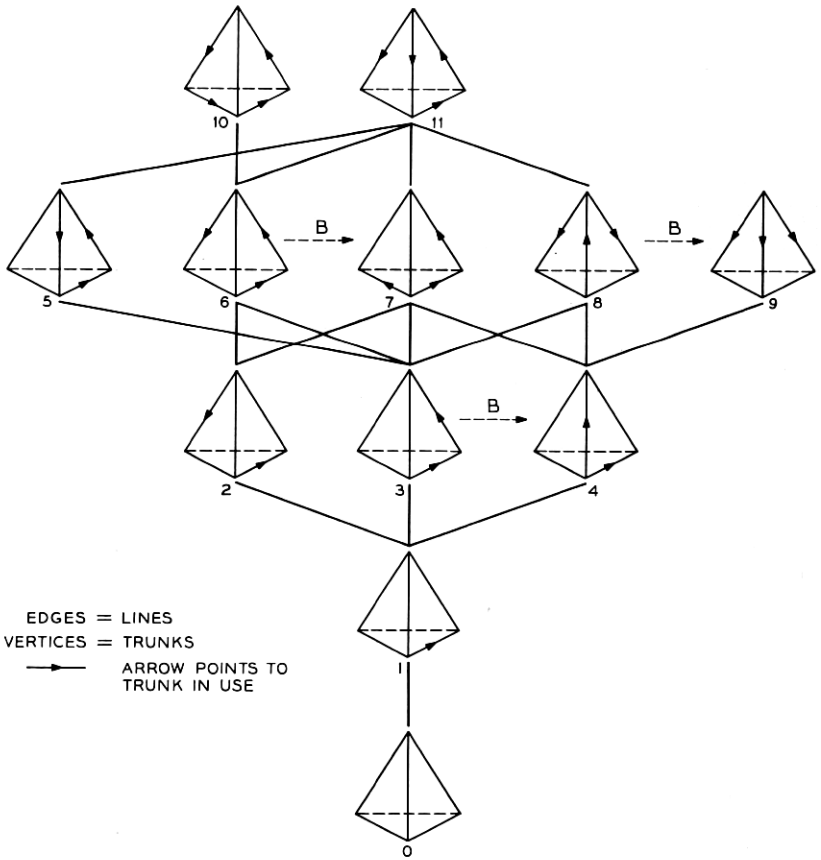

Fig. 4 — States of 6-to-4, 2 access concentrator.

are the key to choosing optimal routes for accepted calls. The resulting optimal policies are remarkably easy to describe and to instrument; they agree fully with some of the conjectures developed over years of practical experience in telephony; they are even robust under changes of performance index. Naturally, each example studied here involves a very small network. Nevertheless, the fact that the monotone properties turned up in each of a substantial number of small networks of diverse structure suggests that they are also present in larger ones. Whether this is so is a topic for future research. In any case, the examples we offer indicate that the theory of routing here developed applies

equally well to central office networks and to various gradings and concentrators.

(*v*) In the interesting area of low traffic, optimal routing can be as much as an order of magnitude better than random routing; with high traffic the advantage decreases rapidly. In all the examples studied, the optimal routing policy was independent of the traffic λ; this suggests that in most cases the optimal policy is basically a combinatorial feature of the network alone, and is probably optimal in many probabilistic models of network operation.

(*vi*) There are situations in which attempted calls should be rejected even though they are not blocked. Simple examples of this phenomenon all seem to be rather unnatural; but J. H. Weber[3] has discovered it numerically in trunking networks, and has suggested[4] that it is associated with unequal lengths of paths for calls. The examples we studied numerically in the present work did not show it; but they had the property that all paths for calls were of the same length. We conjecture that there is a large class of "regular, well-behaved, normal, etc." networks in which no optimal policy rejects an unblocked call, and that in general occasions on which such calls should be rejected are rare. Even if they occur in practical central office networks, these occasions

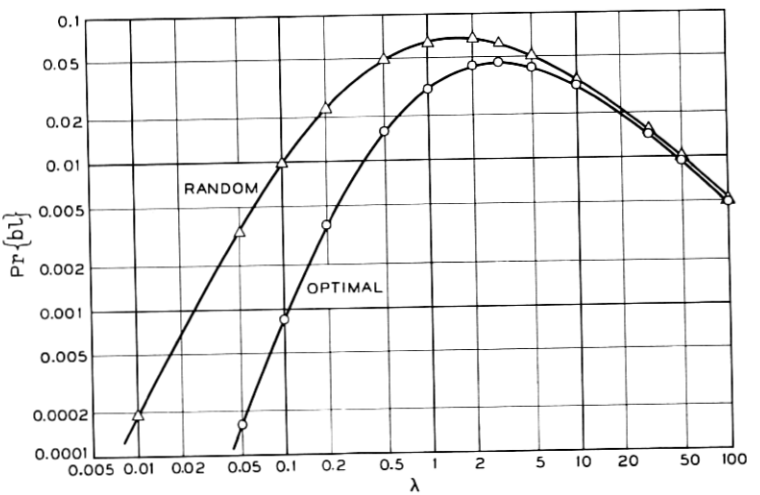

Fig. 5 — $Pr\{bl\}$ for Clos 3-stage network with 2×2 switches.

TABLE I — PROBABILITY OF BLOCKING FOR CLOS 3-STAGE 2 × 2 NETWORK FOR OPTIMAL AND RANDOM ROUTING

λ	$Pr\{bl\}$	
	Optimal	Random
0.01	0.00000181	0.00018319
0.05	0.00015926	0.00334468
0.1	0.00087324	0.00960844
0.2	0.00376107	0.02259477
0.5	0.01593861	0.04807122
1.0	0.03146853	0.06360424
2.0	0.04381783	0.06670098
3.0	0.04584041	0.06206897
5.0	0.04233249	0.05152606
10.0	0.03115608	0.03463135
30.0	0.01405820	0.01459520
50.0	0.00901346	0.00922144
100.0	0.00475109	0.00480733

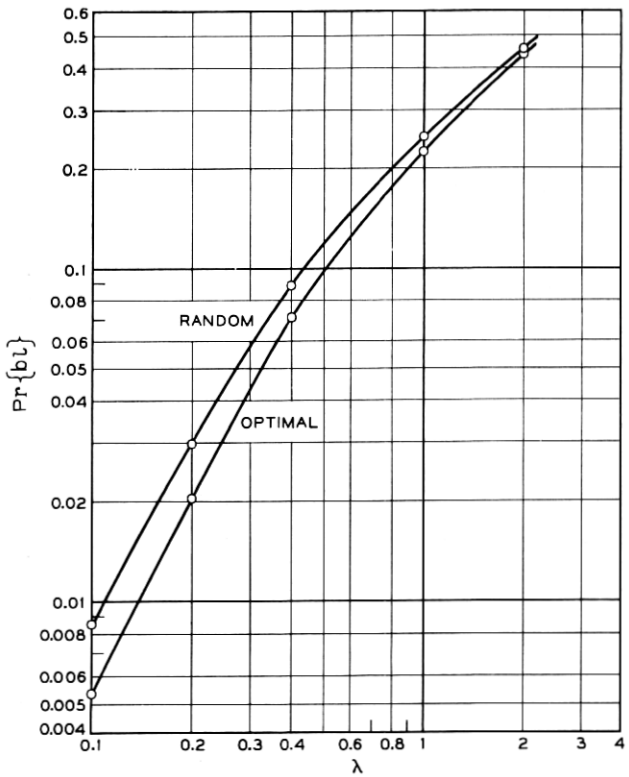

Fig. 6 — $Pr\{bl\}$ for 6-to-4, 2 access concentrator for random and optimal routing.

TABLE II — PROBABILITY OF BLOCKING FOR 6-TO-4, 2 ACCESS CONCENTRATOR FOR OPTIMAL AND RANDOM ROUTING

λ	Optimal		Random
	(S.P. Lloyd)	(Author)	
0.1	0.0049	0.00536231	0.00864729
0.2		0.02093718	0.02972292
0.4	0.0716	0.07170622	0.08856109
0.7	0.1628		
1.0	0.2478	0.23154056	0.24943320
2.0	0.4498	0.44971622	0.46141067

probably should be taken seriously (by a company committed to giving service) only if they are demonstrably associated with large amounts of congestion or a near-breakdown in operation. Hence, finding optimal policies to within rejection of calls may be considered a "practical" solution of the routing problem originally posed.

X. SOME COMPARISON THEOREMS FOR LOW TRAFFIC

There are two ways in which a theoretical analysis can substantially further progress in the problem of routing: (*i*) by means of *local* comparison theorems that establish that one method of routing is better than another, and (*ii*) by means of *global* optimality theorems that exhibit (in part or overall) one or more optimal policies which actually achieve the best possible value of the performance index in use. In this section, we prove some comparison theorems which are valid asymptotically as the traffic parameter approaches zero. At first, we restrict the analysis of the present section to the case[1,2] in which no unblocked call is rejected if it is attempted, so that we avoid the difficult question of deciding whether an attempted call that is not blocked should be completed or rejected.

For a first glimmer of insight, we shall examine the formula

$$Pr\{bl\} = \frac{p'\beta}{p'\alpha}, \qquad Q = Q(R), \qquad R \text{ a fixed rule}$$

valid when no unblocked call is rejected, in the very common situation in which there is an integer greater than zero, n say, such that there is no blocking in states with fewer than n calls in progress, and there are states with n calls in progress in which some calls are blocked. In this case it is known[1,2] that

$$p'\beta = p_0 \frac{\lambda^n}{n!} \sum_{|x|=n} r_x \beta_x + o(\lambda^n),$$
$$p'\alpha = p_0 \sum_{k=0}^{n} \frac{\lambda^k}{k!} \prod_{j=0}^{k-1} \alpha_j + o(\lambda^n), \tag{3}$$

as $\lambda \to 0$, where[1,2]

$$\begin{aligned} r_x &= \text{number of paths on } S \text{ ascending from 0 to } x \text{ and permitted by } R \\ &= (R^{|x|})_{0x} \\ &= \text{the 0,}x \text{ entry of the } |\, x \,| \text{-th power of } R, \end{aligned} \tag{4}$$

and

$$\alpha_j = \text{number of idle inlet-outlet pairs in a state having } j \text{ calls in progress.}$$

(We recall that for the important cases of one- or two-sided networks $\alpha_x = \alpha_{|x|} = \alpha_j$ for all x with $|\, x \,| = j$.) It follows from (3) that for small λ the leading term is critical: the blocking will depend principally on how easy it is to reach a blocking state from the zero state, with this "ease" measured by the number

$$\sum_{|x|=n} r_x \beta_x = (R^n \beta)_0$$

$$= \text{the number of ways in which a blocked call can arise without having any hangups, starting at zero.}$$

If the matrix R is not fixed, but allows some random choices of route, then this quantity can still be viewed as the "expected number of ways in which a blocked call can arise without having any hangups, starting at zero." It is apparent that this number is given by f_0, where the numbers $\{f_x, |\, x \,| \leqq n\}$ are defined by the nonlinear recurrence

$$f_x = \begin{cases} \beta_x & |\, x \,| = n \\ \displaystyle\sum_{\substack{c \in x \\ c \text{ not blocked in } x}} \min_{y \in A_{cx}} f_y & |\, x \,| < n. \end{cases}$$

Indeed we have the result:

Lemma 1:

$$\sum_{|x|=n} r_x \beta_x \geqq f_0 \quad \textit{for} \quad R \in C$$

Proof: Let R be given and let

$$d_x = \begin{cases} \beta_x & |x| = n \\ \sum_{y \in A_x} r_{xy} d_y & |x| < n. \end{cases} \tag{5}$$

We prove the stronger result that $d_x \geqq f_x$. It is clear that

$$d_0 = \sum_{|x|=n} r_x \beta_x, \qquad d_x = f_x \quad \text{for} \quad |x| = n.$$

If $d_y \geqq f_y$ for $|y| = k + 1$, then for $|x| = k$

$$d_x = \sum_{y \in A_x} r_{xy} d_y \geqq \sum_{y \in A_x} r_{xy} f_y$$

$$\geqq \sum_{\substack{c \text{ idle in } x \\ c \text{ not blocked in } x}} \min_{y \in A_{cy}} f_y = f_x.$$

We shall say that $R \in C$ puts $x \in S$ on an ascending path to a state z if and only if $\exists y_0, \cdots, y_{|z|}$ with $y_0 = 0$, $|y_i| = i$, $y_{|z|} = z$, and $r_{y_i y_{i+1}} = 1$ for $i = 0, \cdots, |z| - 1$, and $x > 0$ is among $y_1, \cdots, y_{|z|}$. Let D be the subset of all fixed rules $R \in C$ such that if $|z| = n$, and if R puts x,y with $y \in A_x$ on an ascending path to z, then $r_{xy} = 1$ only if, with $c = \gamma(y - x)$,

$$f_y = \min_{w \in A_{cx}} f_w.$$

The numbers $\{f_x, |x| \leqq n\}$ are the key to optimal routing for low values of λ, or to put it more picturesquely, they are the key to staying as far away as possible from the blocking states in $\{x: |x| = n\}$, which are the ones that provide the leading term in $Pr\{bl\}$ as $\lambda \to 0$. We have

Theorem 1: Let $R \in D$ and $R^* \in C - D$. Then for all λ small enough

$$Pr\{bl\}_R < Pr\{bl\}_{R^*}.$$

Proof: Let d_x^* be defined in terms of R^* according to (5) used in Lemma 1. Since $R^* \notin D$, there exist x,y,c, and $\varepsilon > 0$, such that

$$y \in A_x, \quad \gamma(y - x) = c, \quad r_{xy}^* = 1$$

$$f_y \geqq \min_{z \in A_{cx}} f_z + \varepsilon, \tag{6}$$

and a maximal chain $0 = y_0, y_1, y_2, \cdots, y_{|x|-1}, y_{|x|} = x$ ascending in $\leqq$ such that

$$r_{y_i y_{i+1}}^* = 1. \qquad i = 0, \cdots, |x| - 1.$$

Now, using $d^* \geqq f$,

$$\begin{aligned} d_x^* &= \sum_{z \in A_x} r_{xz}^* d_z^* \\ &= \sum_{A_x - \{y\}} r_{xz}^* d_z^* + f_y \\ &\geqq f_x + \varepsilon, \end{aligned}$$

the last inequality a consequence of $d^* \geqq f$ and the definition of f. Similarly, if $d_{y_{i+1}}^* > f_{y_{i+1}}$, then

$$\begin{aligned} d_{y_i}^* &= \sum_{A_{y_i} - \{y_{i+1}\}} r_{y_i z}^* d_z^* + d_{y_{i+1}}^* \\ &> f_{y_i}^*. \end{aligned}$$

Since $y_0 = 0$, we have $d_0^* > f_0$.

Setting $a = f_0$, $a^* = d_0^*$, and

$$b = \sum_{k=0}^{n} \frac{\lambda^{k-n}}{k!} \prod_{j=0}^{k-1} \alpha_j,$$

we have the asymptotic forms

$$Pr\{bl\}_R = \frac{a + \varepsilon}{b + \delta}$$

$$Pr\{bl\}_{R^*} = \frac{a^* + \varepsilon^*}{b + \delta^*}$$

with $\varepsilon, \delta, \varepsilon^*, \delta^*$ all $o(1)$ as $\lambda \to 0$, and $a < a^*$. Since b increases as $\lambda \to 0$

$$(a - a^* + \varepsilon - \varepsilon^*)b < a^*\delta - a\delta^* + \varepsilon^*\delta - \varepsilon\delta^*$$

for all λ small enough. This is equivalent to

$$ab + \varepsilon b + a\delta^* + \varepsilon\delta^* < a^*b + \varepsilon^*b + a^*\delta + \varepsilon^*\delta,$$

$$\frac{a + \varepsilon}{b + \delta} < \frac{a^* + \varepsilon^*}{b + \delta^*},$$

and proves the theorem.

Low traffic analyses of the kind just employed can also shed some light on the problem of rejecting or accepting unblocked calls. For example, if a call c is *always* refused in every state, then

$$r_{xx} \geqq 1$$

and

$$Pr\{bl\} = \frac{p'(\beta + r)}{p'\alpha} \geqq \frac{1 + p'\beta}{p'\alpha}$$

$$\to \frac{1}{\alpha_0} \quad \text{as} \quad \lambda \to 0.$$

However, if no unblocked call is rejected, then $Pr\{bl\} \to 0$ as $\lambda \to 0$. Thus, *always refusing* c cannot be optimal if λ is sufficiently small.

For another example, suppose as before that

$$n = \min_{y \in S} \{|\, y \,|: \ \beta_y > 0\} > 0,$$

and let c be a call which is refused by R in some state x with $|\, x \,| < n$. It is easy to see that for the rule R

$$Pr\{bl\} \geqq \frac{\dfrac{\lambda^{|x|}}{|x|!} r_x + o(\lambda^{|x|})}{\alpha_0 + o(\lambda)}.$$

On the other hand, if the rule R_1 refuses no unblocked calls,

$$Pr\{bl\} = \frac{\dfrac{\lambda^n}{n!} \sum_{|y|=n} r_y^{(1)} \beta_y + o(\lambda^n)}{\alpha_0 + o(\lambda)},$$

where the superscript 1 indicates that R_1 supplants R in (4). For λ small enough, then

$$\frac{\lambda^{|x|}}{|x|!} r_x > \frac{\lambda^n}{n!} \sum_{|y|=n} r_y^{(1)} \beta_y$$

and R_1 is better than R. Thus, there is never any point in refusing an unblocked call attempt made in a state x whose norm or dimension is less than the minimum norm achieved by the blocking states, if λ is small enough.

XI. REDUCTIONS TO LINEAR PROGRAMMING PROBLEMS

Our effort to choose, with full information about the state of the network, routes for new calls so as to minimize the probability of blocking has led, upon the assumption of a simple probabilistic description for the traffic, to this problem of mathematical programming: To minimize

$$\frac{p'(\beta + r)}{p'\alpha} \tag{7}$$

subject to $p \geqq 0$, $p'1 = 1$, $p'Q = 0$, $Q = Q(R)$, $R \in C$.

It is relatively easy to see that this problem can be formulated as one that has a bilinear (or linear fractional) objective function, and linear constraints. We change variables to $U = (u_{xy})$ and u_{cx} defined by

$$u_{xy} = p_x r_{xy} \qquad x,y \in S, \quad y \in A_x$$

$$u_{cx} = p_x - \sum_{y \in A_{cx}} u_{xy} \qquad c \in x, \quad A_{cx} \neq \theta,$$

$$u_{xx} = \sum_{\substack{c \in x \\ c \text{ not blocked in } x}} u_{cx}.$$

Conversely, we introduce p in terms of U by setting

$$p_x = \begin{cases} \dfrac{\lambda}{|x|} \displaystyle\sum_{y \in B_x} u_{yx} & \text{if} \quad s(x) = 0, \\ \dfrac{u_{xx} + \displaystyle\sum_{y \in A_x} u_{xy}}{s(x)} & \text{if} \quad s(x) > 0. \end{cases}$$

If c is a call which can be completed in state x, then $A_{cx} \neq \theta$, and

$$\lambda \sum_{y \in A_{cx}} u_{xy}$$

is the equilibrium rate at which c is completed in state x, and

$$\lambda u_{cx} = \lambda p_x - \lambda \sum_{y \in A_{cx}} u_{xy}$$

is the equilibrium rate at which c is rejected in state x.

The transformation of variables from p to $\{U,u_{cx}\}$ necessitates adding additional constraints if a sensible problem is to result. Evidently, for $c \in x$ not blocked in x

$$p_x = u_{cx} + \sum_{y \in A_{cx}} u_{xy}.$$

The left-hand side does not depend on c. For different $c \in x$ not blocked in x all these formally different ways of calculating p_x must agree, and it is, therefore, necessary to impose the additional constraint that

$$c,c' \in \gamma(A_{cx} - x) = \gamma\{y: \quad y = z - x \quad \text{for} \quad z \in A_{cx}\} \text{ implies}$$

$$u_{cx} + \sum_{y \in A_{cx}} u_{xy} = u_{c'x} + \sum_{y \in A_{c'x}} u_{xy}.$$

The condition $p'Q = 0$ then gives the condition, for $s(x) > 0$,

$$\frac{|x|}{s(x)} \left(u_{xx} + \sum_{y \in A_x} u_{xy}\right) + \sum_{y \in Ax} u_{xy} = \sum_{y \in A_x} \left(u_{yy} + \sum_{z \in A_y} u_{yz}\right) + \lambda \sum_{y \in B_x} u_{yx}$$

to be satisfied by U. Naturally, the condition $U \geqq 0$ is imposed. We define R in terms of U by

$$r_{xy} = \begin{cases} 0 & \text{unless} \quad y \in A_x \quad \text{or} \quad y = x, \\ \dfrac{u_{xy}}{u_{xx} + \sum\limits_{z \in A_x} u_{yz}} & \text{if} \quad y \in A_x, \\ s(x) - \sum\limits_{y \in A_x} r_{xy} & \text{if} \quad y = x. \end{cases}$$

The normalization condition $p'1 = 1$, finally, amounts in terms of U to

$$\lambda \sum_{s(x)=0} \frac{1}{|x|} \sum_{y \in B_x} u_{yx} + \sum_{s(x)>0} \frac{\left(u_{xx} + \sum\limits_{y \in A_x} u_{xy}\right)}{s(x)} = 1.$$

In terms of U the objective function is

$$\frac{\lambda \sum\limits_{s(x)=0} \frac{\beta_x}{|x|} \sum\limits_{y \in B_x} u_{yx} + \sum\limits_{s(x)>0} \frac{\beta_x}{s(x)} \left(u_{xx} + \sum\limits_{y \in A_x} u_{xy}\right) + u_{xx}}{\lambda \sum\limits_{s(x)=0} \frac{\alpha_x}{|x|} \sum\limits_{y \in B_x} u_{yx} + \sum\limits_{s(x)>0} \frac{\alpha_x}{s(x)} \left(u_{xx} + \sum\limits_{y \in A_x} u_{xy}\right)}.$$

It is possible to describe *linear* programming problems which are equivalent to our nonlinear problem of optimal routing. Two ways of reducing (7) to a linear programming problem will now be discussed. The first is due to A. Charnes and W. W. Cooper.[8] Let $q = tp$, where the scalar $t \geqq 0$ is to be chosen so that $q'\alpha = a$, with $a > 0$ a specified real number. Consider now the "adjoined" *linear* programming problem of finding q,t,r minimizing $q'(\beta + r)$, subject to $q,t \geqq 0$, $q'1 - t = 0$, $q'Q = 0$, $q'\alpha = a$, $Q = Q(R)$, $r = r(R) = \{r_{xx}, x \in S\}$, $R \in C$. (The argument just described shows that the constraints are linear.)

Theorem 2: For any $a > 0$, if q,t,r is a solution of the "adjoined" linear problem, then $p = q/t$ is a solution of (7).

Proof: It is necessary to show first that indeed $t > 0$. Suppose $q,0,r$ is a solution. Then $q'1 = 0$ and $q \geqq 0$ imply $q = 0$, so that $q'\alpha = 0$; but $q'\alpha = a > 0$. Hence, $t > 0$.

If $p'Q = 0$ and $Q = Q(R)$, we use r_p to mean the vector $\{r_{xx}, x \in S\}$. Now suppose that there is a solution p of (7) for which

$$\frac{p'(\beta + r_p)}{p'\alpha} > \frac{q'(\beta + r)}{q'\alpha} = \frac{q'(\beta + r)}{a}. \tag{8}$$

Now $p'\alpha > 0$, because for any $R \in C$ the corresponding value of p_0

(0 = zero state, with no calls up) is > 0, and $\alpha_0 > 0$. Hence, there is a $\theta > 0$ such that $p'\alpha = \theta a$. Consider $\hat{q} = \theta^{-1}p$, $\hat{t} = \theta^{-1}$. Then

$$\theta^{-1}p'\alpha = \hat{q}'\alpha = a$$

and $\hat{q},\hat{t}$ satisfy $\hat{q},\hat{t} \geqq 0$, $\hat{q}'Q = 0$, $\hat{q}'1 - \hat{t} = 0$. But,

$$\frac{p'(\beta + r_p)}{p'\alpha} = \frac{\theta^{-1}p'(\beta + r_p)}{\theta^{-1}p'\alpha} = \frac{\hat{q}'(\beta + r)}{\hat{q}'\alpha} = \frac{\hat{q}'(\beta + r)}{a}.$$

Hence, (8) implies $\hat{q}'(\beta + r_p) > q'(\beta + r)$, because $a > 0$. This contradicts the optimality of q,t,r for the "adjoined" problem.

A cognate reduction to a linear programming problem can be obtained from a lemma of C. Derman,[9] included for completeness:

Lemma 2: The nonlinear function

$$g(x) = \frac{c'x}{d'x}$$

can be minimized subject to $x \geqq 0$, $Ax = b$, *by solving a linear programming problem if* (*i*) $Ax = 0$, $x \geqq 0$ *imply* $x = 0$ *and* (*ii*) $x \geqq 0$, $Ax = b$ *imply* $d'x > 0$.

Proof: Conditions (*i*) and (*ii*) imply that the transformation

$$z = \begin{pmatrix} \dfrac{x}{d'x} \\ \dfrac{1}{d'x} \end{pmatrix}$$

is one-to-one between $\{x \geqq 0, Ax = b\}$ and z satisfying $z \geqq 0$, $d'z = 1$. and $Bz = 0$, where

$$B = (Ab).$$

Under the transformation $g(x)$ becomes a linear function. It can be verified that (*i*) and (*ii*) of Derman's lemma apply to the routing problem (7).

XII. REFORMULATION AS A MARKOV DECISION PROCESS

In Section VII the problem of optimal routing was cast as that of minimizing the probability of blocking, a *bilinear* or *linear fractional* functional of the equilibrium probability vector p, subject to linear constraints. In Section XI it was shown how this problem could be reduced to a linear programming problem which, however, is at best only sug-

gestive in identifying optimal policies. We shall now state an elementary probabilistic result which implies that minimizing the probability of blocking, and maximizing the fraction of events that are successful attempts, are equivalent. This fact permits a direct dynamic programming approach through Markov decision processes, and again leads to a linear programming problem, with the difference, though, that it actually enables us to study optimal policies for many cases, to be described.

Theorem 3: Let p be an equilibrium probability vector for a process x_t resulting from use of some rule $R \in C$. Let

$$m = \sum_{x \in S} |x| p_x = \textit{average number of calls in progress}$$

then both

$$1 - Pr\{bl\} = \frac{1}{1 + \lambda \dfrac{p'(\beta + r)}{m}},$$

and

$$\textit{Fraction of events that are successful attempts} = \frac{1}{2 + \lambda \dfrac{p'(\beta + r)}{m}}.$$

Proof: For the first formula with $s = \{s(x), x \in S\}$

$$Pr\{bl\} = \frac{p'(\beta + r)}{p'\alpha} = \frac{p'(\beta + r)}{p'(s - r) + p'(\beta + r)}$$

and $\lambda p'(s - r) = m$, since the average rate of successes must equal that of hangups, in equilibrium, and $\alpha = \beta + s$.

The second quantity is

$$\frac{\text{average rate of successes}}{\text{average rate of events}} = \frac{\lambda p'(s - r)}{m + \lambda p'\alpha} = \frac{m}{2m + \lambda p'(\beta + r)}.$$

An immediate consequence is:

Theorem 4: Maximizing the fraction of events that are successful attempts is equivalent to minimizing the probability of blocking.

The value of the preceding observations is that we can now reformulate the routing problem as an effort to maximize

$$\lim_{n \to \infty} \frac{1}{n} E\{\text{number of successful attempts in } n \text{ events}\},$$

the asymptotic rate of successful attempts when time is counted discretely, by events.

Since only events are at issue, and the epochs at which they occur are irrelevant, we can discard the continuous parameter Markov process $\{x_t, t \text{ real}\}$ in favor of a Markov chain $\{x_n, n \text{ an integer}\}$, with a transition matrix $A = (a_{xy}) = A(R)$ given by

$$[\,|x| + \lambda\alpha_x]a_{xy} = \begin{cases} \lambda(\beta_x + r_{xx}) & x = y, \\ 1 & y \in B_x, \\ \lambda r_{xy} & y \in A_x, \\ 0 & \text{otherwise.} \end{cases}$$

The stationary vector q satisfying $q = q'A$ is related to p by

$$p_x = (\text{constant})\,\frac{q_x}{|x| + \lambda\alpha_x}.$$

Then

$$E\{\text{number of successful attempts in } n \text{ events}\} = \sum_{j=0}^{n-1} A^j v$$

where $A = A(R)$ and $v = v(R)$ given by

$$\begin{aligned} v_x &= \frac{\lambda s(x) - \lambda r_{xx}}{|x| + \lambda\alpha_x}, \\ &= \text{chance that first event to occur starting in } x \text{ is a successful call.} \end{aligned} \tag{9}$$

Thus, the problem of optimal routing can be cast in the form of the Markov decision processes studied by e.g., R. Bellman[10] and R. Howard:[11] For $R \in C$ and $A = A(R) = (a_{xy})$ given by

$$(|x| + \lambda\alpha_x)a_{xy} = \begin{cases} \lambda(\beta_x + r_{xx}) & x = y, \\ 1 & y \in B_x, \\ \lambda r_{xy} & y \in A_x, \\ 0 & \text{otherwise,} \end{cases}$$

the minimum

$$\min_{R \in C} Pr\{bl\} = \min_{R \in C} \frac{p'(\beta + r)}{p'\alpha}$$

subject to $p'Q = 0$, $p \geqq 0$, $p'1 = 1$ is achieved by the R which maximizes the scalar ρ such that

$$\rho 1 = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} A^j v \quad v = v(R), \quad A = A(R), \quad R \in C$$

with v given by (9).

The results of Bellman in Ref. 10 were derived under the strong positivity condition $a_{xy} \geqq d > 0$ on the matrices A; this condition is of course not met in our routing problem, since many a_{xy} necessarily vanish. However, since our matrices have only one ergodic set it is still possible to obtain results like Bellman's provided only that a little care is taken with the transient states.

Lemma 3: Let ρ be the scalar defined by

$$\rho 1 = \max_{R \in C} \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} A^j v, \tag{10}$$

let R^+ achieve the maximum in (10), and let g be the vector determined[11] *up to a multiple of (the vector) 1 by the equation*

$$\rho 1 + g = v(R^+) + A(R^+)g.$$

Let R^* achieve the maximum in

$$\max_{R \in C} \{v(R) + A(R)g\}$$

Let F be the transient set of states relative to $A(R^*)$. Then the restriction of g to $S - F$ satisfies the nonlinear equation

$$\rho + g_x = \max_{R \in C} \{v_x(R) + \sum a_{xy}(R) g_y\}, \qquad x \in S - F, \tag{11}$$

and the right-hand side of (11) depends in fact only on

$$\{g_y, y \in S - F\}.$$

Further, there is a fixed routing matrix R^{**}, agreeing with R^* on $(S - F)^2$, and a vector g^* agreeing with g on $S - F$, such that R^{**} achieves the maximum in

$$\rho 1 + g^* = \max_{R \in C} \{v(R) + A(R)g^*\}.$$

Proof: If the nonlinear equation given does not hold for some $x \in S - F$.

there exists a vector ζ with $\zeta \neq 0$, $\zeta \geqq 0$ such that on $S - F$

$$\rho + \zeta_x + g_x = \max_{R \in C} \{v_x(R) + \sum_y a_{xy}(R) g_y\},$$
$$= v_x(R^*) + \sum_y a_{xy}(R^*) g_y .$$

Let us restrict all vectors to the $| S - F |$ components present in $S - F$, and the matrix $A(R^*)$ to $(S - F)^2$. Then, dropping dependence on R^*

$$\rho 1 + \zeta + g = v + Ag.$$

There exists an integer k such that $A^k > 0$ strictly. Left-multiply by A^k and note that $A1 = 1$ to obtain

$$\rho 1 + A^k(\zeta + g) = A^k v + A^{k+1} g.$$

Since A^k is a positive matrix, and $\zeta \neq 0$, $\zeta \geqq 0$, there exists a scalar ε such that $A^k \zeta \geqq \varepsilon 1$, so that

$$(\rho + \varepsilon)1 + A^k g \leqq A^k v + A^{k+1} g.$$

Iterating this inequality n times we obtain

$$n(\rho + \varepsilon)1 + A^k g \leqq \sum_{i=k}^{k+n-1} A^i v + A^{k+n} g.$$

For n large enough this contradicts the maximal character of ρ. To find R^{**} and g^*, consider the equation

$$g_x^* = -\rho + \max_{R \in C} \{v_x(R) + \sum_{y \in F} a_{xy}(R) g_y^* + \sum_{y \in S-F} a_{xy}(R) g_y\}, \quad x \in F.$$

This represents the expected best possible fortune of a gambler who starts broke in state $x \in F$, plays by choosing a matrix R paying an amount ρ to play, receiving $v_x(R)$ if he plays R in state x, and ending the game with a final payoff of g_y if the system leaves F for the first time by going into $y \in S - F$; i.e., if he passes through $x_1 x_2 \cdots x_n y$ playing $R_1 R_2 \cdots R_n$ (with R_i in x_i), going out to $y \in S - F$ from x_n, then he receives (or owes)

$$-n\rho + \sum_{i=1}^{n} v_{x_i}(R_i) + g_y .$$

It is apparent that $\{g_x^*, x \in F\}$ exist; R^{**} on $F^2 \cup (F \times S - F)$ is determined by the property that it achieves the maximum above, and on $(S - F) \times F$ it is zero.

Lemma 4: Let ρ be the scalar defined by the condition

$$\rho 1 = \max_{R \in C} \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} A^j v, \tag{12}$$

and let the vector g be a solution of the nonlinear inequality

$$\rho 1 + g \leqq \max_{R \in C} \{v(R) + A(R)g\}. \tag{13}$$

If $R^* \in C$ achieves the maximum on the right of (13), then it also achieves that on the right of (12).

Proof: R^* and g are related by

$$\rho 1 + g \leqq v(R^*) + A(R^*)g,$$

whence, left-multiplying by $A^j = A^j(R^*)$ and summing on j from 0 to $(n-1)$,

$$n\rho 1 + \sum_{j=0}^{n-1} A^j g \leqq \sum_{j=0}^{n-1} A^j v + \sum_{j=1}^{n} A^j g$$

$$\rho 1 - \frac{1}{n} \sum_{j=0}^{n-1} A^j v \leqq o(1) \qquad A = A(R^*), \qquad v = v(R^*).$$

This implies that R^* achieves the maximum in (12).

XIII. OPTIMALITY OF FIXED RULES

If a routing matrix has any entries other than integers, its use introduces a certain amount of additional randomness into the operation of the network, over and above that due to the random traffic, and may be said to represent a "mixed" strategy. It is a natural intuition that since minimizing the probability of loss is a game played against nature, rather than against an intelligent adversary, there can be no real gain from this additional randomization, i.e., that a fixed rule can be found that is as good as any "mixed strategy". To this effect we formulate

Theorem 5: A fixed rule R achieves

$$min \frac{p'(\beta + r)}{p'\alpha}$$

subject to $R \in C$, $p'Q = 0$, $p'1 = 1$, $p \geqq 0$, $Q = Q(R)$.

This theorem is a consequence of the next two results, which, though they are adapted from work of C. Derman,[9] are included here for completeness.

Lemma 5: Let $\xi(\cdot)$: $C \to E^{|S|}$ be an affine map of C into $|S|$-dimen-

sional Euclidean space, i.e., one such that for real scalars $a_1, a_2 \geqq 0$ *with* $a_1 + a_2 = 1$, *and* $R_1, R_2 \in C$,

$$\xi(a_1R_1 + a_2R_2) = a_1\xi(R_1) + a_2\xi(R_2),$$

and let ξ *be continuous. Then,*

$$min\ q'\xi$$

subject to $q \geqq 0$, $q'1 = 1$, $q'A = q$, $A = A(R)$, $\xi = \xi(R)$ *is achieved by a fixed rule* R.

Proof: For $R \in C$ and $A = A(R)$, $\xi = \xi(R)$ set

$$v(R) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} A^j\xi.$$

By a known Markov chain limit theorem,[12] $v(R)$ is well-defined. For $\mu \in (0,1)$ let

$$V(R,\mu) = \sum_{j=0}^{\infty} (\mu A)^j\xi.$$

It is clear that for each $\mu \in (0,1)$, and each starting state x, there exists an $R_{\mu x} \in C$

$$V_x(R_{\mu x}, \mu) = \min_{R\in C} V_x(R,\mu).$$

Then

$$V_x(R_{\mu x}, \mu) = \min_{R\in C} \{\xi_x(R) + \mu \sum_{y\in S} a_{xy}(R)V_y(R_{\mu x}, \mu)\}.$$

The right-hand side is an affine functional of R and so assumes a minimum at an extreme point of C, i.e., at a fixed rule R. Thus, we can consider that $R_{\mu x}$ is a fixed rule. Since the fixed rules form a *finite* class, there exists a sequence $\mu_n \to 1$ and a fixed rule R^* such that

$$R_{\mu_n x} = R^* \qquad n = 1,2, \cdots.$$

By a well-known Abelian theorem,[13] for $R \in C$

$$\lim_{\mu\to 1} (1 - \mu)V(R,\mu) = v(R)$$

and also

$$\begin{aligned} v(R) &\geqq \lim_{n\to\infty} (1 - \mu_n)V(R,\mu_n) \\ &\geqq \lim_{n\to\infty} (1 - \mu_n)V(R^*,\mu_n) \\ &\geqq v(R^*). \end{aligned}$$

Thus, R^* is optimal.

Theorem 6: Let $\zeta,\eta: C \to E^{|S|}$ *be affine maps of* C *into* $|S|$*-dimensional Euclidean space, and let* ξ *and* η *be continuous, with* $\eta(R) > 0$ *for* $R \in C$. *Then*

$$b = \min \frac{q'\zeta}{q'\eta}$$

subject to $q \geqq 0$, $q'A = q$, $q'1 = 1$, $A = A(R)$, $\xi = \xi(R)$, *and* $\eta = \eta(R)$ *is achieved by a fixed rule.*

Proof: Let $b(R)$ be the value of $q'\xi/q'\eta$ for a given choice R, with q determined by the constraints $q \geqq 0$, $q'A = q$, $q'1 = 1$. There exist $R_1, R_2, \cdots \in C$ such that

$$\lim_{n\to\infty} b(R_n) = b.$$

For n fixed, let $\xi(\cdot)$ in Lemma 5 be given by

$$\xi = \zeta - b(R_n)\eta.$$

Then in the notation of Lemma 5, $v(R_n) = 0$. By Lemma 5 there exists a fixed rule R_n^* such that

$$\begin{aligned} v(R_n^*) &\leqq v(R_n) \\ &\leqq 0, \end{aligned}$$

that is, since $q'\eta \neq 0$,

$$b(R_n^*) \leqq b(R_n).$$

Since there is a finite number of fixed rules, there is a subsequence $n_1, n_2, \cdots$ and a fixed rule R^* such that $R_{n_i}^* = R^*$, $i = 1, 2, \cdots$. Then R^* is optimal.

XIV. TRYING TO GET CLOSER TO THE OPTIMAL ROUTING RULES

It is particularly important to try to verbalize, and eventually to mechanize, routing strategies that are optimal, near-optimal, or by some yardstick just "good". In this endeavor, the fact that the original routing problem (7) can be formulated and solved numerically as a linear programming problem, while interesting theoretically and perhaps reassuring, is nevertheless of limited usefulness. For this reason we have attempted to take advantage of some of the special properties of the problem that are due to its telephonic origins, and to describe at least parts of optimal policies in terms of the combinatorial properties of the connecting network upon which they ultimately depend.

In the second half of this paper we introduce some additional notions and assumptions of a combinatorial nature. With their aid we are able to exhibit parts of some actual optimal routing rules. The problem of finding out something concrete about optimal policies has been so difficult that we have quite frankly started with (and so far restricted attention to) cases which can be treated by what T. M. Burford has called "domination" arguments, which depend on or establish isotony[5] properties for certain networks having suitable monotone structures. The word 'monotone' is used loosely here: more specifically, the networks are to have the property that the relative merit of states is consistent or continuous, i.e., that if one state x is "better" than an equivalent state y, then the *neighbors* of x are in the same sense "better" than the corresponding neighbors of y.

Although some of the combinatorial properties (on which the results to be given are based) are strong, we believe that these properties and the optimal policies (or partial policies) they lead to have a definite relevance to the practical aspects of optimal routing, if only because they bear out some of the intuitive conjectures offered in Section VIII. Our results show not only that these conjectures are "in the right ballpark," but also that in many instances they describe optimal policies.

We start our discussion with four simple examples; once the ideas involved are understood, the principles behind them can be abstracted, and general theorems proved.

It has been shown (Section XII) that minimizing the probability of blocking is equivalent to maximizing the fraction of events that are successful attempts, where an event is either a hangup, a blocked attempt, or a successful one. This maximal fraction is the limit, as n becomes large, of

$$\frac{1}{n} E_x(n),$$

where

$E_x(n)$ = expected number of successful calls in n events, if the network starts in state x and an optimal policy is followed.†

We shall base our approach on the vectors $E(n)$.

First example: Consider the overflow system or grading shown in Figs. 7 and 8. There are two groups of lines, each of two lines; the first has access to both trunks to the destination, but the second has access to the second trunk only. The possible states of this system (reduced under the

† Here an optimal policy is one for which the expected number of successful calls in n steps is a maximum.

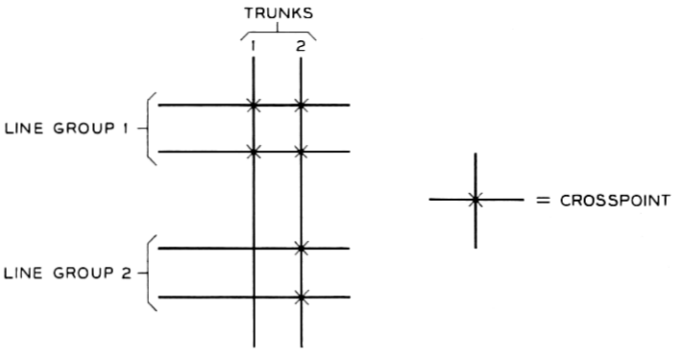

Fig. 7 — Asymmetric grading.

equivalence relation induced by permuting lines within a line group) form the partially ordered system of Fig. 8. There is only one situation which demands a choice between alternative routes for a call; it arises when a call from line group 1 is accepted with no calls in progress. The two alternatives are indicated in Fig. 8 by the notation "ch": one is to put the call on trunk 1, leaving no lines blocked, the other is to put it on trunk 2, leaving 2 lines blocked.

What circumstances make one choice of a route better than another? In the present instance it is clear that use of trunk 1 for a group 1 call in state 0 leaves the "high access" trunk 2 free to serve group 2. Thus, at first glance a route whose use blocked the smallest possible number of additional calls (over and above those that are already blocked) seems to be best. It is natural to expect that in state 0 a new call from group 1 should be routed on trunk 1 and not on trunk 2. Indeed, it can be shown that if such a call should be accepted then it should be placed on trunk 1. (For small λ it *should* always be accepted, as was proved in Section X.) Thus, a policy which routes a group 1 call on trunk 1 in state 0 can differ

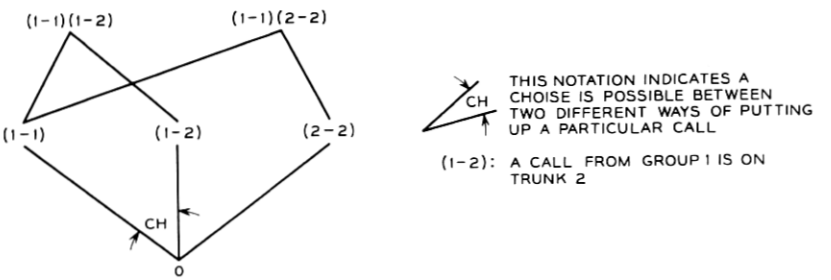

Fig. 8 — States of the grading of Fig. 7.

from an optimal policy only in that it might accept some calls which the other rejected, and *vice versa*.

Rather than proving the result stated above, we shall discuss other examples, involving different kinds of network: it will turn out that similar circumstances arise. Indeed, we shall claim that the particular circumstance on which the result is based is no isolated happenstance, but a phenomenon common enough to be relevant to the theory of routing. All examples discussed here, as well as many others, will be covered by a general result (Theorem 14) proved later.

Second example: Referring to Fig. 2, which shows the reduced state diagram of the three-stage Clos network of Fig. 1, we observe that only in the state numered 4 are there any blocked calls. State 4 realizes the same assignment of inlets to outlets as state 2, which has no blocked calls. The difference between the two is that in state 2 all the traffic passes through one middle switch, leaving the other entirely free for any call that may arise. This difference illustrates the intuitive rule that one should always put a call through the most heavily loaded part of the network that will still accept it. This example was discussed in Refs. 1, 2 where it was shown (rather laboriously) that if no calls are rejected, then preferring state 2 to state 4 in state 1 is optimal. This result will be an instance of Theorem 14.

Third example: It is to be expected that in some instances a choice of route for a call is immaterial. The concentrating switch depicted in Figs. 9 and 10 is a simple example of this phenomenon. It is intuitively obvious that, because of the symmetries of the network, it makes no difference which of the two trunks a call could use when the system is empty is assigned to it. This insensitivity of performance to routing choices can actually be deduced from Theorem 7.

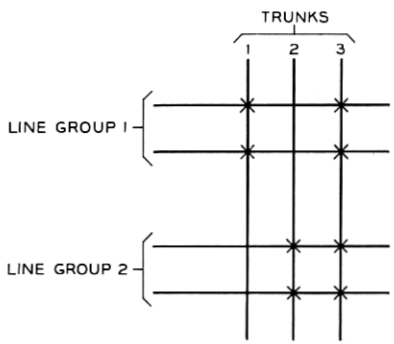

Fig. 9 — Symmetric grading.

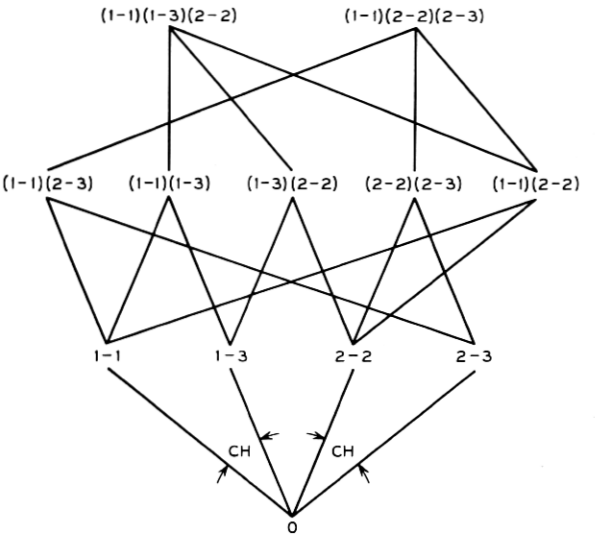

Fig. 10 — States of the grading of Fig. 9.

Fourth example: Figs. 11 and 12 show the structure and (reduced) state diagram for another simple Clos network made of 3 × 3 inlet and outlet switches, and 2 × 2 middle switches. Again, from scrutiny of the state diagram we guess that optimal routing will result if no empty middle switches are used when partially filled ones are available. The notations '*B*' in Fig. 12, intended to suggest that the states to the left of the *B*'s are "better" than those on the right, constitute an expression of the corresponding policy, and are explained in the next paragraphs.

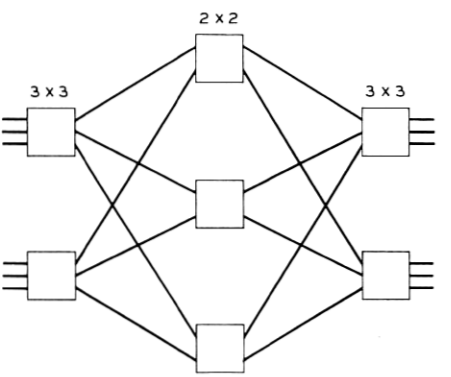

Fig. 11 — 3-stage Clos network with 3 × 3 outer switches.

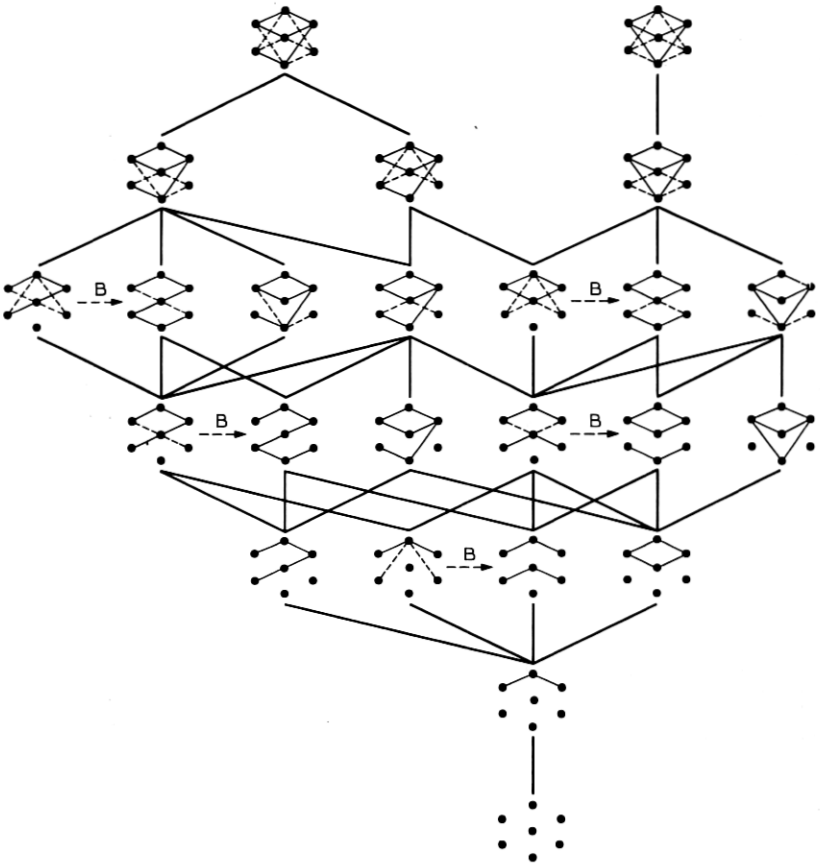

Fig. 12 — States of 3-stage Clos network of Fig. 11.

To abstract the essential features of the preceding examples into a general theorem, we start with the observation that in choosing to enter a state x rather than another y in putting up a call we have always to choose between *equivalent* states ($x \sim y$, in the sense of Section III), in which the same events e can occur. In particular, the same new calls c can arise. If it now happens that every new call blocked in x is also blocked in y, let us regard this as *prima facie* evidence that x is somehow "better" than y, and define a relation $B \subseteq S^2$ by the condition

$$xBy \text{ if and only if } x \sim y \text{ and}$$

$$c \in x,\ c \text{ blocked in } x \quad \text{imply} \quad c \text{ blocked in } y.$$

The relation B is a partial ordering.

In the first example considered above, (1-1)B(1-2), and B obtains between no other distinct states; in the second, 2B4, and again B obtains between no other distinct states.

Let us now suppose (for a general network with state set S) that the network is run according to a policy φ, and ask what happens to B under φ. That is, more specifically, we look at states x,y such that xBy, and we consider, for events e that are either hangups or new calls blocked in neither x nor y, whether or not

$$\varphi(e,x)B\varphi(e,y).$$

If e occurs and φ is used for decisions, then the system moves from x to $\varphi(e,x)$ and from y to $\varphi(e,y)$. If $\varphi(e,x)B\varphi(e,y)$ for all $e \in x$ that are either hangups or new calls blocked in neither x nor y, whenever xBy, we say that φ *preserves* B. Formally,

φ preserves B if and only if xBy implies $\varphi(e,x)B\varphi(e,y)$ for $e \in x$ which are either hangups or new calls blocked in neither x nor y.

In the first example (Fig. 8) there are no new calls c which can be put up in both (1-1) and (1-2), and there is one hangup (say h) which can occur in both. Thus, the set of events to be considered is just $\{h\}$. Clearly, $\varphi(h,1\text{-}1) = \varphi(h,1\text{-}2) = 0$ state for any φ. Since B is reflexive, we conclude that in this case every φ preserves B.

In the second example, a similar situation arises. There are two events to be considered: one is a new call completable in both 2 and 4 leading to state 6, the other is a hangup leading to 1. Again

$$\varphi(e,2) = \varphi(e,4)$$

for all φ and both events e to be considered, and again any φ preserves B.

As noted, routing has no effect in the third example. However, the relation B is defined. It can be verified that any φ preserves B, and that in this case B is a symmetric relation, as it should be, since if routing is to have no effect, then x can only be "just as good" as y if y is "just as good" as x. These facts can be used to *prove* that routing has no effect in this example.

The fourth example, finally, shows the relation B in action. The notations

$$x - - B - \rightarrow y \quad x,y \text{ states}$$

in Fig. 12 show the *irreflexive* part of B. (Obviously xBx for all $x \in S$, and this part of B is not shown in Fig. 12.) The reader is invited to

verify that the policy φ of using a partly-filled middle switch whenever possible does indeed preserve B in this example.

The property of a policy φ, that it preserves B, is to be viewed as a kind of *isotony* of φ:

$$xBy \quad \text{implies} \quad \varphi(e,x)B\varphi(e,y), \quad \text{for suitable } e.$$

(See G. Birkhoff,[5] p. 3.) It can also be viewed as a kind of *continuity*, for after all if we think of the set of *neighbors* N_y of y as the states in

$$N_y = A_y \cup B_y,$$

then the property says that if xBy then also zBw where z is a neighbor of x and w a neighbor of y such that $z \sim w$. In other words it states that if xBy then also

$$(N_x \times N_y) \cap (\sim) \subseteq B,$$

i.e., if it holds between x and y then it also holds between equivalent neighbors of x and y.

Note that if φ preserves B, xBy, and φ rejects in x a call c not blocked in y, then it also rejects it in y.

For φ a policy, let

$E_x(n,\varphi)$ = expected number of successful attempts in n events, if the network starts in state x and policy φ is followed.

The isotonic property that φ preserve B has the useful feature that it implies an isotony among the numbers

$$\{E_x(n,\varphi), \quad n \geqq 1, \quad x \in S\}.$$

This is the content of the next result.

Theorem 7: (First Isotony Theorem): If φ preserves B, then xBy implies

$$E_x(n,\varphi) \geqq E_y(n,\varphi), \qquad n = 1, 2, \cdots.$$

Proof: xBy, $c \in x$, $\varphi(c,y) \neq y$ imply $\varphi(c,x) \neq x$. Hence,

$$\sum_{\substack{c \in x \\ \varphi(c,x)=x}} 1 \leqq \sum_{\substack{c \in y \\ \varphi(c,y)=y}} 1,$$

and $E_x(1,\varphi) \geqq E_y(1,\varphi)$. As a hypothesis of induction assume that xBy implies

$$E_x(n,\varphi) \geqq E_y(n,\varphi)$$

for some $n \geqq 1$. We have

$$E_x(n+1,\varphi) = \sum_{\substack{c\in x \\ \varphi(c,x)\neq x}} \frac{\lambda}{|x|+\lambda\alpha_x}\{1+E_{\varphi(c,x)}(n,\varphi)\}$$

$$+\frac{\lambda}{|x|+\lambda\alpha_x}E_x(n,\varphi)\sum_{\substack{c\in x \\ \varphi(c,x)=x}} 1+\frac{1}{|x|+\lambda\alpha_x}\sum_{h\in x}E_{x-h}(n,\varphi).$$

Since φ preserves B, it must be true that xBy implies

$$\varphi(c,x)B\varphi(c,y)$$

$$(x-h)B(y-h),$$

whence

$$E_{\varphi(c,x)}(n,\varphi) \geqq E_{\varphi(c,y)}(n,\varphi)$$

$$E_{x-h}(n,\varphi) \geqq E_{y-h}(n,\varphi).$$

Therefore,

$$E_x(n+1,\varphi) \geqq \sum_{\substack{c\in y \\ \varphi(c,y)\neq x}} \frac{\lambda}{|x|+\lambda\alpha_y}\{1+E_{\varphi(c,y)}(n,\varphi)\}$$

$$+\frac{\lambda}{|x|+\lambda\alpha_y}E_y(n,\varphi)\sum_{\substack{c\in y \\ \varphi(c,y)=y}} 1$$

$$+\frac{1}{|y|+\lambda\alpha_y}\sum_{y\in h}E_{y-h}(n,\varphi)$$

$$\geqq E_y(n+1,\varphi).$$

The power and utility of the relation B are further illustrated by the following comparison theorem for policies. The partial ordering B on S induces a natural partial ordering B of the policies according to the definition

$$\varphi B\psi \equiv e \in x,\ x \in S \quad \text{imply} \quad \varphi(e,x)B\psi(e,x)$$

for e a hangup or a call not blocked in x. We note that $\varphi B\psi$ implies that φ and ψ embody the same rejection policy.

Theorem 8: If $\varphi B\psi$, and one of φ,ψ preserves B, then xBy implies

$$E_x(n,\varphi) \geqq E_x(n,\psi), \qquad n = 1, 2, \cdots.$$

Proof: φ and ψ have the same rejection policy, so $E(1,\varphi) = E(1,\psi)$, and the theorem holds for $n = 1$. Assume as a hypothesis of induction

that xBy implies $E_x(n,\varphi) \geqq E_y(n,\psi)$ for a given value of $n \geqq 1$. We have, with $p_{ex} = \Pr\{e \text{ occurs in } x\}$,

$$E_y(n+1,\varphi) = E_y(1,\varphi) + \sum_{e \in y} p_{ey} E_{\varphi(e,y)}(n,\varphi).$$

But $e \in y$ implies $\varphi(e,y) B \psi(e,y)$, and so by the induction hypothesis

$$E_{\psi(e,y)}(n,\varphi) \geqq E_{\psi(e,y)}(n,\psi).$$

However,

$$\begin{aligned} E_y(n+1,\psi) &= E_y(1,\psi) + \sum_{e \in y} p_{ey} E_{\psi(e,y)}(n,\psi) \\ &\leqq E_y(n+1,\varphi). \end{aligned}$$

Let now xBy, and suppose that φ preserves B. The isotony theorem then implies

$$\begin{aligned} E_x(n+1,\varphi) &\geqq E_y(n+1,\varphi) \\ &\geqq E_y(n+1,\psi). \end{aligned}$$

If, instead, ψ preserves B, then

$$E_x(n+1,\psi) \geqq E_y(n+1,\psi)$$

and a repetition of the first part of the argument above with x instead of y gives

$$\begin{aligned} E_x(n+1,\varphi) &\geqq E_x(n+1,\psi) \\ &\geqq E_y(n+1,\psi). \end{aligned}$$

XV. SECOND INTUITIVE APPROACH

In an effort to develop a more general theory than the one that was begun in the previous two sections, we now make a fresh start at understanding the structure of "good" routing; again, we begin with a special case:

Fifth example: We choose the overflow system or grading depicted in Fig. 13. There are two groups of lines, one of two lines, the other of three lines. Each has access to one primary trunk to which the other does not have access, and they share a single common overflow trunk. The possible states of this system form the partially ordered system shown in Fig. 14. Alternative ways of putting up particular calls are marked with "ch", for "choice".

After inspecting the system and its state diagram, intuition tells us

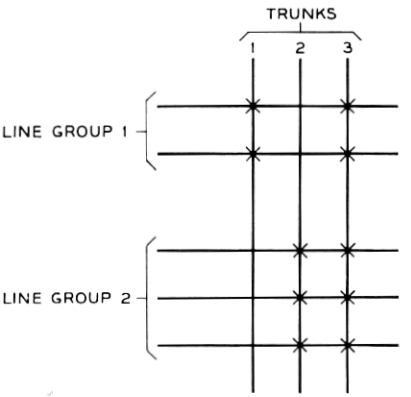

Fig. 13 — Second asymmetric grading.

that, as a first guess, calls should use the primary trunks whenever they can, so as to leave the overflow open as much as possible. Let us, on this basis, formulate some preferences for certain routes.

Clearly, in state 0 a call from group 1 should go on trunk 1, so in state 0 we prefer state (1-1) to (1-3); similarly we prefer (2-2) to (2-3). The same principle should apply if certain calls are already in progress. Thus, in state (2-2) we prefer (1-1) (2-2) over (1-3) (2-2), and in state (1-1) we prefer (1-1) (2-2) to (1-1) (2-3).

If taken seriously and followed, the preferences listed above define a

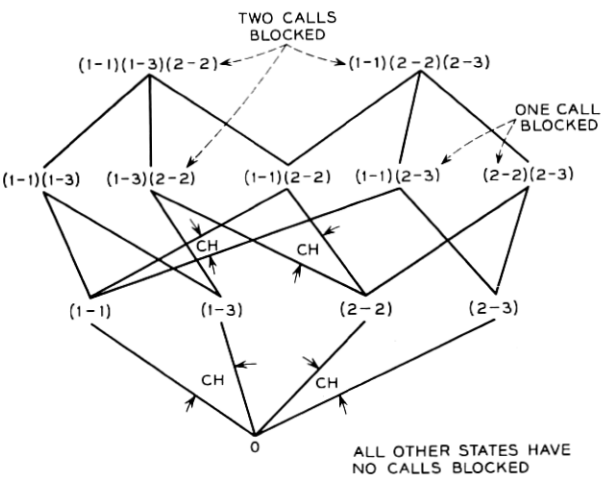

Fig. 14 — States of the grading of Fig. 13.

policy for putting in calls. We shall show that this policy differs from the optimal policy only in that the latter may reject some calls, while the former accepts all unblocked calls. To do this write xPy if state x is preferred to state y. Thus, the relation P is defined by the conditions

$$(1\text{-}1)\ P\ (1\text{-}3)$$

$$(2\text{-}2)\ P\ (2\text{-}3)$$

$$(1\text{-}1)\,(2\text{-}2)\ P\ (1\text{-}3)\,(2\text{-}2)$$

$$(1\text{-}1)\,(2\text{-}2)\ P\ (1\text{-}1)\,(2\text{-}3).$$

We let

$E_x(n)$ = expected number of successful call attempts in n events, if the system starts in state x and an optimal policy is used.

It must be explained here that by "use of an optimal policy" over n steps we mean simply that we use a policy which will maximize the average number of successful attempts *among those n events*; the policies that achieve this may, for all we know at this point, be different for different n.

A slight departure from the probabilistic model of Section VI is necessary here: we assume that an idle line generates calls to the trunk destination at a rate $\lambda > 0$, instead of assuming that an idle inlet-outlet pair generates calls at λ. Also, we let α_x be the number of idle lines in x, rather that than that of idle inlet-outlet pairs, and $s(x)$ that of idle lines that are not blocked.

Theorem 9: If xPy, then

$$E_x(n) \geqq E_y(n) \qquad n = 1, 2, 3, \cdots .$$

Proof:

$$E_x(1) = \frac{\lambda s(x)}{|x| + \lambda\alpha_x},$$

and xPy implies $s(x) \geqq s(y)$, so the theorem is true for $n = 1$. Assume that the theorem holds for some $n \geqq 1$. There are four cases, corresponding to the four conditions defining P. We shall give the argument for the case where

$$x = (1\text{-}1)\,(2\text{-}2)$$

$$y = (1\text{-}3)\,(2\text{-}2),$$

and (as we know) xPy; the others are similar.

Now apparently

$$E_{(1\text{-}1)(2\text{-}2)}(n+1) = \frac{1}{2+3\lambda}\{E_{(2\text{-}2)}(n) + E_{(1\text{-}1)}(n)\}$$
$$+ \frac{\lambda}{2+3\lambda}\max\{E_{(1\text{-}1)(2\text{-}2)}(n),\ 1 + E_{(1\text{-}1)(1\text{-}3)(2\text{-}2)}(n)\}$$
$$+ \frac{2\lambda}{2+3\lambda}\max\{E_{(1\text{-}1)(2\text{-}2)}(n),\ 1 + E_{(1\text{-}1)(2\text{-}2)(2\text{-}3)}(n)\}$$

and

$$E_{(1\text{-}3)(2\text{-}2)}(n+1) = \frac{1}{2+3\lambda}\{E_{(2\text{-}2)}(n) + E_{(1\text{-}3)}(n)\}$$
$$+ \frac{\lambda}{2+3\lambda}\max\{E_{(1\text{-}3)(2\text{-}2)}(n),\ 1 + E_{(1\text{-}1)(1\text{-}3)(2\text{-}2)}(n)\}$$
$$+ \frac{2\lambda}{2+3\lambda}E_{(1\text{-}3)(2\text{-}2)}(n).$$

By the induction hypothesis,

$$E_{(1\text{-}1)}(n) \geqq E_{(1\text{-}3)}(n)$$
$$E_{(1\text{-}1)(2\text{-}2)}(n) \geqq E_{(1\text{-}3)(2\text{-}2)}(n);$$

hence,

$$E_x(n+1) \geqq E_y(n+1)$$

for the given x and y.

The point is that each event that can occur leads to a "worse" state in y than it does in x. Thus, the hangup of the group 1 call leads both to the state 2-2, a standoff; hangup of the group 2 call takes x into (1-1) and y into (1-3), and (1-1)P(1-3); one of the possible new calls leads both x and y to the state (1-1)(1-3)(2-2), another standoff; the other two possible new calls are blocked in y but not in x, so that by the induction hypothesis, rejecting one of them and staying in x is at least as good as having one of these blocked calls make an attempt in y.

We conclude from Theorem 9 that in an optimal policy the calls which are not rejected are put on the primary trunks if these are available, and on the overflow only if the primary trunk appropriate to the call is already busy. This result is entirely in agreement with our original intuition.

Another example of the same kind is shown in Figs. 15 and 16: the intuitive preferences shown in Fig. 16 by 'P' are optimal to within rejection of unblocked calls.

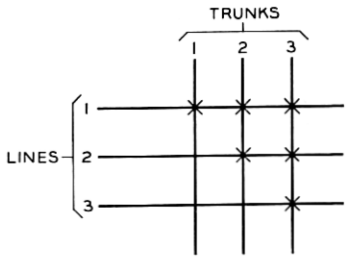

Fig. 15 — Third asymmetric grading.

We now formalize the principles behind the intuitions that led to Theorem 9.

Let P be a relation on S, i.e., a subset of S^2. We may as well put our cards on the table and indicate that P is to be interpreted as a relation of "preference", with xPy meaning "x is preferred to y". If μ is a function, and X,Y are sets, the (customary) notation

$$\mu: \quad X \leftrightarrow Y$$

means that μ takes X into Y in a one-one manner, while

$$\mu: \quad X \to Y$$

means that the μ-image of X is contained in y.

We say that P has the *strong monotone property* if xPy implies

(*i*) $|\, x \,| = |\, y \,|$

(*ii*) $\exists\mu$: $B_x \leftrightarrow B_y$ such that $z \in B_x$ implies $zP\mu z$

(*iii*) $\exists\nu$: $A_y \to A_x$ such that

$$\begin{aligned} &\nu(A_{cy}) \subseteq A_{cx} \quad \text{for} \quad c \in y, \\ &z \in A_y \quad \text{implies} \quad \nu z P z. \end{aligned} \tag{14}$$

Let us denote by F_x the set of all calls which are *free* or idle in x, i.e.

$$\begin{aligned} F_x &= \{c: \quad c \text{ is idle in } x\} = \{\gamma(y - x): \quad y \in A_x\} \\ &= \{c: \quad c = \{(u,v)\} \subseteq I \times \Omega \text{ with } u,v \text{ both idle in } x\}. \end{aligned}$$

We say that a relation P on S has the *weak monotone property* if xPy implies

(*i*) $|\, x \,| = |\, y \,|$

(*ii*) $\exists\mu$: $B_x \leftrightarrow B_y$ and $z \in B_x$ implies $zP\mu z$

(*iii*) $\exists\nu$: $F_y \to F_x$ and $c \in F_y$, $z \in A_{cy}$

$$\text{imply } \exists w \in A_{(\nu c)x} \quad \text{with} \quad wPz. \tag{15}$$

To get the weak monotone property from the strong, define ν on F_y by

$$\nu\gamma(z - y) = \gamma(\nu z - x), \qquad z \in A_y;$$

then $z \in A_{cy}$ implies $\nu z \in A_x$, and

$$\nu c = \gamma(\nu z - x);$$

thus,

$$\nu z \in A_{(\nu c)x} \quad \text{and} \quad \nu z P z.$$

Keeping in mind the interpretation that 'xPy' means that x is in some sense better than y, we see that: condition (i) restricts P to hold only between states of the same norm or dimension, because we are interested only in choosing between states with the same number of calls in progress; condition (ii) says roughly that to every hangup leading out of state y there corresponds a hangup in x leading to a state which is at least as "good" (as the one reached by the hangup in y); condition (iii) says that for any way of completing a new call c in y there is a way of completing the *same* call c in x which leads to at last as "good" a state (as the one reached by completing that call in y).

It is easily seen that P has one of the monotone properties if and only if xPy implies that P holds between "corresponding respective" neigh-

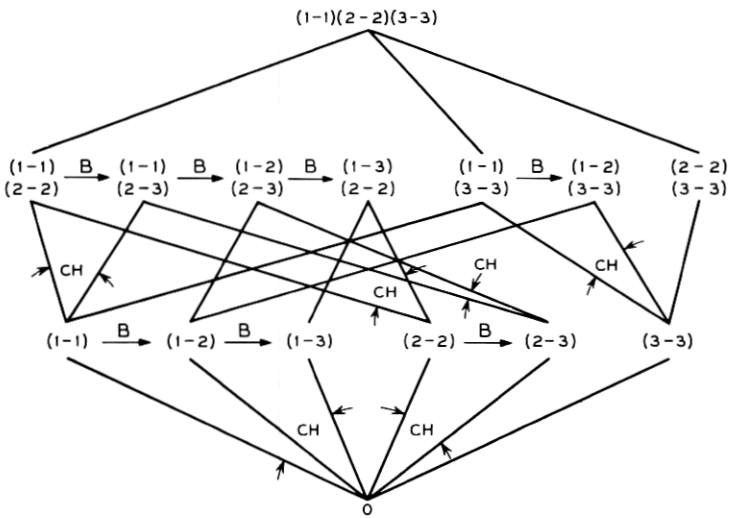

Fig. 16 — States of the grading of Fig. 15.

bors of x and y. Thus, the monotone properties are similar to the property of a policy φ that it preserve B. The principal differences are that here no policy is at issue, and that the meaning of "corresponding neighbor" is weaker than in the definition of preservation. The relationships to the relation B are further clarified in the following remarks.

If P has the weak monotone property, then xPy implies $s(x) \geqq s(y)$. If P has the strong monotone property, then xPy implies that every $c \in x$ blocked in x is blocked in y. Further, since we are primarily interested in comparing *equivalent* states (i.e., x and y such that $x \sim y$), it is natural to restrict attention to preference relations P which are subsets of $\sim$, $P \subseteq \sim$. It can then be verified that if P has either monotone property, and holds only between equivalent states ($P \subseteq \sim$), then $P \subseteq B$.

A "preference" relation should impose at least a partial ordering among the objects for which it is defined, and so it is by nature transitive. The question then arises whether the relations P that have the (strong or weak) monotone property are reflexive and transitive. It is obvious that if P has the monotone property then so does $I \cup P$ where I is the identity relation. Now, as is known, every relation P can be extended uniquely to its *transitive closure* $\bar{P}$, the smallest transitive relation containing P. We shall now prove:

Theorem 10: If $P \subseteq S^2$ has the weak monotone property, then so does its transitive closure $\bar{P}$.

Proof: Clearly $\bar{P} = P \cup P^2 \cup P^3 \cup \cdots$, where the powers represent relative, not Cartesian, products. It is obvious that $x\bar{P}y$ implies $|x| = |y|$, so $\bar{P}$ has property (i) of (14). Next let $x\bar{P}y$, so that there exist $z_1, z_2, \cdots, z_n \in S$ such that $z_1 = x$, $z_n = y$ and

$$z_i P z_{i+1} \qquad i = 1, \cdots, n-1.$$

Thus, there exist maps $\mu_1, \mu_2, \cdots, \mu_{n-1}$ with $\mu_i : B_{z_i} \leftrightarrow B_{z_{i+1}}$ such that $z \in B_{z_i}$ implies

$$zP\mu_i z.$$

Hence, $z \in B_x$ implies

$$zP\mu_1 z$$

$$\mu_1 zP\mu_2\mu_1 z$$

$$\vdots$$

$$\mu_{n-2}\mu_{n-3} \cdots \mu_1 zP\mu_{n-1}\mu_{n-2} \cdots \mu_1 z,$$

i.e.,

$$z\bar{P}\left(\prod_{j=1}^{n-1}\mu_j\right)z.$$

Thus,

$$\mu = \prod_{j=1}^{n-1}\mu_j$$

has the property that μ: $B_x \leftrightarrow B_y$ and $z \in B_x$ implies $z\bar{P}\mu z$. Hence, $\bar{P}$ has property (*ii*). Finally, there exist maps $\nu_1, \cdots, \nu_{n-1}$ with ν_{n-i}: $F_{z_{i+1}} \to F_{z_i}$ such that $c \in F_{z_{i+1}}$, $z \in A_{cz_{i+1}}$ implies $w \in A_{(\nu_i c)z_{i+1}}$ with wPz. Let

$$\nu = \prod_{i=1}^{n-1}\nu_i .$$

Hence, for each $c \in F_y$, $w_n \in A_{cy}$ there exist $w_{n-1}, \cdots, w_n \in S$ and $c_{n-1}\cdots c_{n-1}$such that

$$c_i = \nu_i c_{i+1},\ w_i \in A_{c_i z_i}, \qquad w_i P w_{i+1} \qquad i = 1, \cdots, n-1.$$

It is apparent that $c_1 = \nu c$, $w_1 \in A_{(\nu c)x}$ and $w_i\bar{P}w_n$, so that $\bar{P}$ has property (*iii*).

The following result is now immediate:

Theorem 11: If P has the weak monotone property, and I is the identity relation, then

$$\overline{(I \cup P)}$$

is a partial ordering relation with the weak monotone property.

Any relation with the weak monotone property can be extended to be a *partial ordering* P that has the weak monotone property. Since $\sim$ is an equivalence relation between states, and P is a partial ordering, it follows that $P \cap \sim$ is also a partial ordering.

Theorem 12: (Second Isotony Theorem): If $P \subseteq S^2$ has the weak monotone property, then

$$xPy \quad \textit{implies} \quad E_x(n) \geqq E_y(n), \qquad n = 1, 2, \cdots .$$

Proof: Property (15) (*iii*) implies that $s(x) \geqq s(y)$ whenever xPy. Now

$$E_x(1) = \frac{\lambda s(x)}{|x| + \lambda\alpha_x}.$$

Since it is assumed that $\alpha_x = \alpha_{|x|}$ we have, by (15) (i),

$$xPy \quad \text{implies} \quad E_x(1) \geqq E_y(1).$$

As an hypothesis of induction assume that xPy implies $E_x(n) \geqq E_y(n)$. We have

$$E_x(n+1) = \frac{1}{|x| + \lambda\alpha_x} \sum_{c \in x} \max \{E_x(n), g(c,x) + \max_{z \in A_{cx}} E_z(n)\} + \frac{\lambda}{|x| + \lambda\alpha_x} \sum_{h \in x} E_{x-h}(n),$$

and a similar expression for $E_y(n+1)$. If now xPy, then $|x| = |y|$ by (15) (i), and also

$$E_{x-h}(n) \geqq E_{\mu(x-h)}(n)$$

by (15) (ii) and the hypothesis of induction. Similarly,

$$\frac{\lambda\beta_y}{|x| + \lambda\alpha_x} E_x(n) \geqq \frac{\lambda\beta_y}{|y| + \lambda\alpha_y} E_y(n).$$

For c not blocked in y, and $z \in A_{cy}$, xPy implies that there exists $w \in A_{(\nu c)x}$ with wPz, by (15) (iii). By the hypothesis of induction, this implies that

$$E_w(n) \geqq E_z(n).$$

Since $z \in A_{cy}$ was arbitrary, we find

$$g(\nu c,x) + \max_{w \in A_{(\nu c)x}} E_w(n) \geqq g(c,y) + \max_{z \in A_{cy}} E_z(n).$$

It follows that xPy implies $E_x(n+1) \geqq E_y(n+1)$.

XVI. RELEVANCE OF THE ISOTONY THEOREMS TO OPTIMAL POLICIES

Let $c \in x$ be a call that is not blocked in state x, so that $A_{cx} \neq \theta$. If the hypotheses of one of the isotony theorems obtain, then it may be possible to single out some of the states $y \in A_{cx}$ as providing ways of completing c in x which are at least as good as certain others. Specifically, the sort of comparison we can make is this: If $y,z \in A_{cx}$ and yBz or yPz, then y is at least as good as z in the sense that

$$E_y(n) \geqq E_z(n), \qquad n = 1, 2, \cdots.$$

Suppose now that there is at least one $y \in A_{cx}$ such that yBz for *all* $z \in A_{cx}$. It then follows that such a y is always at least as good a choice

as any other state of A_{cx}, in the above sense. A similar result follows if there is a $y \in A_{cx}$ with yPz for all $z \in A_{cx}$. In such situations a policy that routes c so as to take the system from x to y can differ (so far as x and c are concerned) from an optimal policy only in the respect that an optimal policy might reject c in x. This is the sense in which the isotony theorems can provide the part of the solution of the routing problem which has to do with choosing routes for accepted calls. Two theorems to this effect appear in Section XVIII after an aside about equivalence of decisions and nonuniqueness of optimal policies.

XVII. EQUIVALENCE OF DECISIONS AND NONUNIQUENESS OF OPTIMAL POLICIES

It is natural to expect that there are often several optimal policies, in the sense that, for some c and x with $c \in x$ and $A_{cx} \neq \theta$, there are two choices of a route for c in x which are in some sense distinct routes and yet are both equally "good". For example, in most traffic models for a graded or progressive multiple it often does not make any difference which trunk in a group is used for a call: the possible states resulting from use of one of the trunks in the group are all distinct, yet all are equally "good", being "equivalent" under permutations of trunks within the group. It is intuitively clear that such a nonuniqueness of optimal policies is due in large part to symmetries in the network under study, or more generally, to the presence of various equivalences of states (and hence of routing decisions) under certain groups of permutations of terminals.† Since some of these equivalences appear in a later proof, we digress a little for an account of them, first heuristic, then formal.

As we have seen, one of the principal tools in the description of optimal policies is a combinatorial partial ordering, such as B or P, which implies an ordering in terms of performance. The discussion to follow is based on a general partial ordering R, which the reader can assume is contained in

$$\bigcup_{c \in x} A_{cx}^2$$

and which he can interpret as B or P, if he wishes.‡

Let then R be a partial ordering of S and let Y be a subset of S. Cued by the remarks of Section XVI, we want to use R to compare states; in

† It should be noted that the word 'group' is used in this paragraph in two technical senses, the first from traffic theory, referring to a set of trunks, the second from the theory of groups.

‡ This use of 'R' is peculiar to this section, and should not be confused with R as a routing matrix.

particular we wish to talk about elements $y \in Y$ such that yRz for all $z \in Y$. It would be satisfyingly simple if at this point we could introduce the notation

$$\sup_R Y$$

for that element of Y which bears R to every other element of Y. Unfortunately this is usually impossible, because there may be several or many such "suprema" of Y. In this situation the usual mathematical trick to use is to pass to suitable equivalence classes. Use of this procedure is further justified by the fortunate fact that, in the case of several interesting choices of R and Y, there are several senses in which these maximal elements turn out to be equivalent. What is more, there is a natural equivalence based only on R, such that $\sup_R Y$ can, if it exists, be defined in the "quotient" set of the equivalence, i.e., in the image of the semilattice homomorphism that takes each state into the equivalence class to which it belongs.

If $R = P$ and P has the monotone property, then all the P-suprema of A_{cx} are equivalent in this very important sense: If $y_1, \cdots, y_m$ is an enumeration of all the $y \in A_{cx}$ that are best in the sense that yPz for all $z \in A_{cx}$, then

$$y_i P y_j, \qquad 1 \leqq i, j \leqq m$$

and the second isotony theorem gives

$$E_{y_i}(n) = E_{y_j}(n) \qquad n = 1, 2, \cdots, \tag{16}$$

so that *as far as performance is concerned*, $y_1, \cdots, y_m$ are all "equivalent". In many cases, this fact is based on an underlying equivalence of a combinatorial nature, much stronger than (16): e.g., in a trunk group the different states attainable by different choices of a trunk for a call are equivalent in the sense that given any two there is a renaming or permutation of the trunks which carries one into the other.

The isotony theorems provide ways of translating a *combinatorial comparison* of states such as

$$xBy, \quad \text{or} \quad xPy$$

into a *numerical comparison* of the relative merit or value of starting in each state, x or y. In such a setting it is natural to call x and y "equivalent" if the comparison holds both ways, i.e., if, when interpreting 'xBy' as a (rather strong) precise form of 'x is better than y', we have both

$$xBy \quad \text{and} \quad yBx.$$

Lemma 6: Given two states y,z there exists at most one pair c,x such that both

$$y,z \in A_{cx} .$$

Proof: If $y,z \in A_{cx}$ then $x = y \cap z$ in the sense of the semi-lattice of states. Thus, x is unique. If now

$$y,z \in A_{cx} \cap A_{c'x}$$

then $c = \gamma(y - x)$, $c' = \gamma(y - x)$, so $c = c'$.

The foregoing observations are the motivation for the ensuing development. With the partial ordering R we associate the natural equivalence relation $\equiv_R$ defined by

$$z \equiv_R y \quad \text{if and only} \quad zRy \quad \text{and} \quad yRz \quad \text{and} \quad \exists A_{cx} \quad y,z \in A_{cx} .$$

The subscript R will usually be dropped as long as it is contextually clear what R is being used to define $\equiv$. Along with $\equiv$ we introduce the semilattice homomorphism

$$r(\cdot)\colon S \to \{\text{equivalence classes of } \equiv\} = S/\equiv$$

$$\text{defined by } \tau(x) = \{z\colon \ z \equiv x\} .$$

The image $\tau(S)$, i.e., the "quotient" set $S/\equiv$, is partially ordered by the relation R defined by

$$\tau(x)R\tau(y) \quad \text{if and only} \quad u,v \quad u \in \tau(x) \quad \text{and} \quad v \in \tau(y) \quad \text{and} \quad uRv.$$

This is the natural homomorphic "contraction" of R to $S/\equiv$. It can be verified that if $\tau(x)R\tau(y)$ and $\tau(y)R\tau(x)$, then $\tau(x) = \tau(y)$ strictly.

If now Y contained in S is such that there exists a $y \in Y$ with yRz for every $z \in Y$, we use the notation

$$\sup_R Y \tag{17}$$

for $\tau(y)$. It is clear that in the "quotient" space, an element maximal with respect to R is unique if it exists at all. Strictly speaking the notation

$$\sup_{\tau R} \tau Y$$

would be better, since it indicates that the supremum operation only makes sense after the homomorphism. However, (17) will be used, with the reminder that it is a set, not a state, and the convention that use of (17) implies the assumed existence of maximal elements.

With the notation (17) we can prove the following natural relation-

ship between the strong monotone property and the notion of preservation of B.

Theorem 13: Let

$$\varphi(e,x)\begin{cases}\in \sup\limits_{B} A_{e,x} & \text{for} \quad e = c \\ = x - h & \text{for} \quad e = h\end{cases}$$

and suppose that φ preserves B. Then B has the strong monotone property.

Proof: xBy implies $x \sim y$ and hence $|x| = |y|$, so B has property (14), (i). If xBy, define for $z \in B_x$

$$\mu z = \varphi(\gamma(x - z),y).$$

Then, since φ preserves B

$$\varphi(\gamma(x - z),x)B\varphi(\gamma(x - z),y),$$

$$zB\mu z,$$

and B has property (14) (ii). With xBy still, let

$$\nu: \quad A_y \to A_x$$

be given by $\nu z = \varphi(c,x)$ for $z \in A_{cy}$. Then, since $\varphi(c,y)Bw$ for $w \in A_{cy}$,

$$\varphi(c,x)B\varphi(c,y)$$

$$Bz,$$

so that B has property (14) (iii).

XVIII. OPTIMAL ROUTING THEOREMS

This final section contains precise statements showing just how the combinatorial properties introduced in Sections XIV and XV answer the question: "Which route should an accepted call use?"

Two policies φ and ψ will be termed *equivalent with respect to rejections*, written $\varphi \sim \psi$, if they both reject the same calls in the same states, i.e., if $\varphi(c,x) = x$ when and only when $\psi(c,x) = x$ for $c \in x$.

Theorem 14: If φ preserves B, and if $c \in x$ implies

$$\varphi(c,x) \in \sup_{B} A_{cx}$$

whenever $\varphi(c,x) \neq x$, then

$$E_x(n,\varphi) \geqq E_x(n,\psi) \qquad n = 1, 2, \cdots.$$

for any $\psi \sim \varphi$.

Proof: $E_x(1,\varphi) = E_x(1,\psi)$ by direct calculation. Assume as a hypothesis of induction that $E_x(n,\varphi) \geqq E_x(n,\psi)$ for $x \in S$. We have

$$E_x(n+1,\varphi) = \sum_{\substack{c \in x \\ \varphi(c,x) \neq x}} \frac{\lambda}{|x| + \lambda\alpha_x} \{1 + E_{\varphi(c,x)}(n,\varphi)\}$$

$$+ \frac{1}{|x| + \lambda\alpha_x} \sum_{\substack{c \in x \\ \varphi(c,x) = x}} E_x(n,\varphi)$$

$$+ \frac{1}{|x| + \lambda\alpha_x} \sum_{h \in x} E_{x-h}(n,\varphi),$$

and a similar expression for $E_x(n+1,\psi)$. If now $\varphi(c,x) \neq x$, then $\varphi(c,x) B y$ for every $y \in A_{cx}$; in particular, $\psi(x,x) \neq x$ because $\varphi \sim \psi$, and so $\psi(c,x) \in A_{cx}$, whence

$$\varphi(c,x) B \psi(c,x).$$

The first isotony theorem and the induction hypothesis now give

$$E_{\varphi(c,x)}(n,\varphi) \geqq E_{\psi(c,x)}(n,\varphi)$$

$$\geqq E_{\psi(c,x)}(n,\psi).$$

It follows that

$$E_x(n+1,\varphi) \geqq E_x(n+1,\psi).$$

Corollary: If φ preserves B, and

$$\varphi(c,x) \in \sup_B A_{cx}$$

for $c \in x$ not blocked in x, then φ is optimal within the class of policies that reject no unblocked calls.

Theorem 15: If P has the weak monotone property, and

$$\sup_P A_{cx}$$

exists for each $c \in x$ not blocked in x, then there exists an optimal policy R such that $c \in x$, $y \in A_{cx}$ imply either x is R-transient or else

$$r_{xy} = 0 \quad \textit{unless} \quad y \in \sup_P A_{cx}.$$

Proof: Let ρ be the scalar such that

$$\rho 1 = \max_{R \in C} \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} A^j v \qquad \begin{cases} v = A(R), \\ A = A(R). \end{cases}$$

We first use an argument of R. Bellman[10] to show that the vector sequence

$$E(n) - n\rho\mathbf{1} = g(n)$$

is bounded in n.

By Lemma 3, there is a vector g^* which satisfies

$$g^* + \rho\mathbf{1} = \max_{R \in C} \{v(R) + A(R)g^*\} .$$

Choose $K > 0$ so that

$$g^* - K\mathbf{1} \leqq g(1) \leqq g^* + K\mathbf{1}.$$

Assume, as an induction hypothesis, that

$$g^* - K\mathbf{1} \leqq g(n) \leqq g^* + K\mathbf{1}.$$

We have

$$g(n+1) = -\rho\mathbf{1} + \max_{R \in C} \{v(R) + A(R)g(n)\} .$$

Hence,

$$-\rho\mathbf{1} - K\mathbf{1} + \max_{R} \{v(R) + A(R)g^*\} \leqq g(n+1)$$

$$\leqq -\rho\mathbf{1} + K\mathbf{1} + \max_{R} \{v(R) + A(R)g^*\}$$

$$g^* - K\mathbf{1} \leqq g(n+1) \leqq g^* + K\mathbf{1}.$$

Let now

$$g = \limsup_{n \to \infty} g(n),$$

taken componentwise. Let R_n achieve the maximum in

$$\max_{R \in C} \{v(R) + A(R)g(n)\} .$$

Given $\varepsilon > 0$, there exists n_0 such that $n > n_0$ implies

$$g_x(n) \leqq g_x + \varepsilon$$

for all $x \in S$. Thus,

$$v(R_n) + A(R_n)g(n) = v(R_n) + A(R_n)g + A(R_n)[g(n) - \bar{g},$$

$$\leqq v(R_n) + A(R_n)g + \varepsilon$$

$$\leqq \max_{R \in C} \{v(R) + A(R)g\} + \varepsilon.$$

Hence, since $\varepsilon > 0$ was arbitrary,

$$g + \rho 1 \leqq \max_{R \in C} \{v(R) + A(R)g\} . \tag{18}$$

Let R^* achieve the maximum on the right above. By Lemma 3, R^* is optimal. Let F be the set of transient states relative to R^*. The argument used in Lemma 4 shows that equality must obtain in (18) on $S - F$, i.e.,

$$g_x + \rho = \max_{R \in C} \left\{ v_x(R) + \sum_{y \in S-F} a_{xy}(R) g_y \right\}, \qquad x \in S - F.$$

This is equivalent to

$$g_x + \rho = \sum_{\substack{c \in x \\ c \text{ not blocked in } x}} \frac{\lambda}{|x| + \lambda \alpha_x} \max \left\{ g_x, 1 + \max_{z \in A_{cx}} g_z \right\}$$

$$+ \frac{\lambda \beta_x g_x}{|x| + \lambda \alpha_x} + \frac{1}{|x| + \lambda \alpha_x} \sum_{h \in x} g_{x-h}, \qquad x \in S - F.$$

Now the second isotony theorem implies that if xPy, then

$$E_x(n) \geqq E_y(n), \qquad n \geqq 1$$

$$g_x(n) \geqq g_y(n), \qquad n \geqq 1$$

$$g_x \geqq g_y .$$

Thus, if $c \in x$ is not blocked in x

$$\max_{z \in A_{cx}} g_z$$

is achieved by each and any $y \in \sup_P A_{cx}$.

Let R be any routing matrix such that for $y \in A_{cx}$

$$r_{xy} = \begin{cases} 0 & \text{if } \ y \in S - F, \\ 1 & \text{only if } \ 1 + g_y \geqq g_x \ \text{ and } \ y \in \sup_P A_{cx} . \end{cases}$$

Then R achieves the maximum in (18), and so is optimal; it is clear that it also has the property claimed in the theorem.

XIX. ACKNOWLEDGMENT

The author wishes to thank J. H. Suurballe and D. Alcalay for assistance in using the LP 90 program, F. Sinden and E. Wolman for numerous discussions, E. Wolman again for a careful reading of the draft, and S. P. Lloyd for access to unpublished manuscripts.

APPENDIX

Expected Number of Events to the First Blocked Call

The purpose of this appendix is to demonstrate that if the index of performance is changed to one which attaches greater importance (than does $Pr\{bl\}$) to blocked calls occurring soon after the system is started, then no unblocked call should ever be rejected. This result can be obtained for various indices of performance; we obtain it for the expected number of events occurring until the first blocked call. This choice of index of performance has a natural heuristic justification: it corresponds to trying *to put off the undesirable event* (a blocked call) as long as possible. (Time is being measured here in discrete units, by counting events.)

As before we use φ and ψ for policies, but here we limit them to *rejection policies*, or policies for the acceptance or rejection of unblocked calls. We may think of φ as a binary function of c,x with $c \in x$ and c not blocked in x, and interpret $\varphi(c,x) = 1$ as acceptance, and $\varphi(c,x) = 0$ as rejection. A general routing policy, such as described by a fixed routing matrix R, will be said to be *within* φ if it accepts and/or rejects the same calls in the same states.

We first introduce the quantities

$E_x(\varphi)$ = Expected number of events until the first blocked or rejected call under a routing policy optimal within the rejection policy φ, starting in x.†

These satisfy the equations

$$E_x(\varphi) = \frac{|x| + \lambda s(x)}{|x| + \lambda\alpha_x} + \frac{\lambda}{|x| + \lambda\alpha_x} \sum_{\substack{c \in x \\ c \text{ not blocked in } x \\ \varphi(c,x) = 1}} \max_{y \in A_{cx}} E_y(\varphi) + \frac{1}{|x| + \lambda\alpha_x} \sum_{h \in x} E_{x-h}(\varphi).$$

Our object will be to pick the best rejection policy, i.e., to choose φ so as to achieve

$$\max_{\varphi} E_x(\varphi).$$

We next define, for each fixed routing matrix R

$E_x(R)$ = Expected number of events until the first blocked or rejected call, starting in x and using the policy R.

† The word 'optimal' here refers, naturally, to the fact that the (not necessarily stationary) policy followed makes the expected number of events to the first call (rather than $Pr\{bl\}$, or some other index) a maximum.

For fixed φ, let $R^* = R^*(\varphi) = (r_{xy}{}^*)$ be a routing matrix with the property

$$r_{xy}{}^* = \begin{cases} 1 & \text{if } c \in x \text{ such that } y \in A_{cx}, \quad \varphi(c,x) = 1, \\ & \text{and } E_y(\varphi) = \max_{z \in A_{cx}} E_z(\varphi), \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that at least one such R^* exists, that it is within φ, and that it defines a stationary policy for which

$$E(R^*) = E(\varphi).$$

We now partially order all rejection policies thus:

$\varphi \geqq \psi$ if and only if $\varphi(c,x) \geqq \psi(c,x)$ for $c \in x$ not blocked in x.

Let $\mathfrak{R}$ be the set of rejection policies. The principal result is that $E(\cdot)$ is *isotone* on the partial ordering $\geqq$ of $\mathfrak{R}$, expressed in

Theorem 16: $\varphi \geqq \psi$ *implies* $E(\varphi) \geqq E(\psi)$.

Proof: For $|S|$-vectors v define the transformations T_φ, $\varphi \in \mathfrak{R}$ by

$$(T_\varphi v)_x = \frac{\lambda}{|x| + \lambda\alpha_x} \sum_{\substack{c \in x \\ \varphi(c,x)=1 \\ c \text{ not blocked in } x}} \max_{y \in A_{cx}} v_y + \frac{1}{|x| + \lambda\alpha_x} \sum_{h \in x} v_{x-h}.$$

With

$$b_x = \frac{|x| + \lambda s(x)}{|x| + \lambda\alpha_x},$$

the equation for $E(\varphi)$ becomes

$$E(\varphi) = b + T_\varphi E(\varphi).$$

It is evident that if $v \geqq 0$ and $\varphi \geqq \psi$, then

$$T_\varphi v \geqq T_\psi v.$$

Furthermore, each T_φ, $\varphi \in \mathfrak{R}$, is a monotone transformation in that

$$v \geqq w \quad \text{implies} \quad T_\varphi v \geqq T_\varphi w.$$

Hence, $v \geqq w \geqq 0$, $\varphi \geqq \psi$ imply

$$b + T_\varphi v \geqq b + T_\psi w.$$

For $\varphi \geqq \psi$, then, consider the rectangular parallelopiped

$$\mathcal{P} = \{v: \ 0 \leqq v \leqq E(\varphi)\}.$$

For $v \in \mathcal{P}$ we have

$$E(\varphi) = b + T_\varphi E(\varphi) \geqq b + T_\psi v,$$

so that $T_\psi : \mathcal{P} \to \mathcal{P}$. It is obvious that $\mathcal{P}$ is closed and that T_ψ is continuous. Hence, by Brouwer's fixed point theorem there is a $v \in \mathcal{P}$ satisfying

$$v = b + T_\psi v.$$

We next show that v is actually the unique solution of this equation, so that $v = E(\psi) \leqq E(\varphi)$. Introduce the norm $\| v \| = \max_{x \in S} v_x$. The case in which the network under study is nonblocking and ψ rejects no calls is trivial. Assume then that there exists a state x and a call $c \in x$ such that either c is blocked in x or c is not blocked in x and is rejected by ψ. This implies that the "matrix" part of T_ψ is strictly substochastic, and hence that for some n

$$\| T_\psi^n \| < 1.$$

Thus, $v = E(\psi)$.

It is an immediate consequence of Theorem 16 that if $\varphi^*(c,x) \equiv 1$ for $c \in x$ not blocked in x, then

$$E(\varphi^*) = \max_{\varphi \in \mathfrak{R}} E(\varphi).$$

REFERENCES

1. Beneš, V. E., Markov Processes Representing Traffic in Connecting Networks, B.S.T.J., *42*, November, 1963, pp. 2795–2838.
2. Beneš, V. E., *Mathematical Theory of Connecting Networks and Telephone Traffic*, Academic Press, New York, 1965.
3. Weber, J. H., Some Traffic Characteristics of Communication Networks with Automatic Alternate Routing, B.S.T.J., *41*, March, 1962, pp. 1201–1247.
4. Weber, J. H., private communication.
5. Birkhoff, G., *Lattice Theory*, Revised Edition, Amer. Math. Soc. Coll. Publ., *XXV*.
6. Ketchledge, R. W., The No. 1 Electronic Switching System, IEEE Trans. Comm. Tech., *COM-13*, pp. 38–41, and references therein.
7. Kalaba, R. and Juncosa, M., Optimal Design and Utilization of Communication Networks, Manage. Sci., *3*, 1956, pp. 33–44.
8. Charnes, A. and Cooper, W. W., Programming with Linear Fractional Functionals, Naval Research Logistics Quarterly, *9*, 1962, pp. 181–185.
9. Derman, C., On Sequential Decisions and Markov Chains, Manage. Sci., *9*, 1962, pp. 16–24.
10. Bellman, R., A Markovian Decision Process, J. Math. Mech., *6*, 1957, pp. 679–684.
11. Howard, R., *Dynamic Programming and Markov Processes*, Technology Press and John Wiley & Sons, New York, 1960, p. 62.
12. Chung, K. L., *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, 1960, p. 32.
13. Widder, D. V., *The Laplace Transform*, Princeton University Press, Princeton, 1946, p. 180.
14. *Ibid.*, p. 181.