Algebraic and Topological Properties of Connecting Networks

By V. E. BENEŠ

(Manuscript received May 29, 1961)

A connecting network *is an arrangement of switches and transmission links allowing a certain set of terminals to be connected together in various combinations, usually by disjoint chains (paths): e.g., a central office, toll center, or military communications system. Some of the basic combinatory properties of connecting networks are studied in the present paper.*

Three of these properties are defined informally as follows: A network is rearrangeable *if, given any set of calls in progress and any pair of idle terminals, the calls can be reassigned new routes (if necessary) so as to make it possible to connect the idle pair. A state of a network is a blocking state if some pair of idle terminals cannot be connected. A network is* nonblocking in the wide sense *if by suitably choosing routes for new calls it is possible to avoid all the blocking states and still satisfy all demands for connection as they arise, without rearranging existing calls. Finally, a network is* nonblocking in the strict sense *if it has no blocking states.*

A distance between states can be defined as the number of calls one would have to add or remove to change one state into the other. This distance defines a topology on the set of states. Also, the states can be partially ordered by inclusion in a natural way. This partial ordering and its dual define two more topologies for the set of states. The three topologies so obtained are used to characterize (i.e., give necessary and sufficient conditions for the truth of) the three properties of rearrangeability, nonblocking in the wide sense, and nonblocking in the strict sense. Each of these three properties represents a degree of abundance of nonblocking states; the mathematical concept used to express these degrees is the topological notion of denseness.

Table of Contents

I. INTRODUCTION

Any large communication system contains a *connecting network*, an arrangement of switches and transmission links through which certain terminals can be connected together in many combinations, usually by many different possible routes through the network. Examples of connecting networks can be found in telephone central offices, toll centers, and military communications systems.

The connections in progress in a connecting network usually do not arise in a predetermined temporal sequence; instead, requests for connection (new calls) and terminations of connection (hangups) occur more or less "at random." For this reason it is customary to use the performance of a connecting network when subjected to random traffic as a figure of merit. One precise measure of this performance is the fraction of requested connections that cannot be completed in a given time interval, or the *probability of blocking*. In a telephone connecting network this probability measures to some extent the grade of service given to the customers.

The performance of a connecting network for a given traffic level is determined largely by its configuration or structure. This configuration may be described by stating what terminals or transmission links have a switch placed between them and can be connected together by closing the switch. The configuration of a connecting network determines what groups of terminals can be connected together simultaneously. Any one set of permissible connections may be called a *state* of the network. Quantities such as the number of combinations of terminals that can be connected, and the number of states in which a given combination is connected, clearly are indicative of both the performance and the cost of the system. If these numbers are small the performance may be poor and the cost low; if large, the performance may be unnecessarily good and the cost prohibitive. These numbers are among the purely combinatory and topological properties of the connecting network.

For example, in a telephone exchange, the network configuration determines what pairs of terminals can be simultaneously connected by disjoint paths, that is, what calls can be in progress. If this configuration is too simple, only a few pairs of terminals can have calls in progress between them at the same time. If the configuration is extensive and complicated it may provide for many large groups of simultaneous calls in progress, but the network itself may be expensive to build and difficult to control.

To design connecting networks with confidence, then, it is desirable to have an adequate general understanding of their combinatory and topological properties. A discussion, in part heuristic and tutorial, of connecting systems and of some associated mathematical problems has been given in a paper[1] by the author; the reader is referred thereto for material suitable as background for the present paper. In that work a division of the topic into *combinatory*, *probabilistic*, and *variational* problems was drawn, and it was argued that the elements of this division had a natural order of priority: one must know the combinatory properties of a system in order to calculate its probabilistic properties, i.e., its performance in the face of random traffic; and one must know both the combinatory and the probabilistic properties of systems in order to compare them and to select optimal ones.

In this paper we shall be concerned exclusively with those combinatory and topological properties of a general connecting network that seem to be most relevant to its performance.

II. SUMMARY

Some of the basic combinatory properties of connecting networks are studied in the present work. Three of these properties, rearrangeability, nonblocking in the wide sense, and nonblocking in the strict sense, can be defined informally as follows: for brevity, define an *idle pair* to be a pair of idle terminals consisting of an inlet and an outlet. A network is *rearrangeable* if, given any set of calls in progress, and any idle pair, the existing calls can be assigned new routes (if necessary) so as to make it possible to connect the idle pair. A state of a network is a *blocking state* if some idle pair cannot be connected. A network is *nonblocking in the wide sense* if by suitably choosing routes for new calls it is possible to avoid all the blocking states and still satisfy all demands for connection as they arise, without having to rearrange existing calls. Finally, a network is *nonblocking in the strict sense* if it has no blocking states.

A distance between states of a connecting network can be defined as the number of pairs of terminals that are connected in one state and not in the other. This distance defines a topology on the set of states. Also, the states can be partially ordered by inclusion in a natural way. This partial ordering and its dual define two more topologies for the set of states. The three topologies so obtained are used to characterize (i.e., give necessary and sufficient conditions for the truth of) the three properties of rearrangeability, nonblocking in the wide sense, and nonblocking in the strict sense. Each of these three properties represents a degree of abundance of states in which calls are not blocked; the mathematical concept used to express these degrees is the topological notion of *denseness*. A study of some particular connecting networks that are rearrangeable is given in another paper.[2]

III. THE STRUCTURE AND CONDITION OF A CONNECTING NETWORK

In discussing connecting networks, we shall abstract from the many possible technological realizations and actual designs of connecting networks, and shall consider only certain relevant features on which we can base a useful and sufficiently general mathematical theory.

Most real telephone switching networks consist of pairs of wires for talking paths and electromechanical switches for crosspoints; in certain experimental systems the talking paths are pulse code modulation channels, and the crosspoints are time-division gates made of transistors. However, any attempt to formulate some general properties of connecting networks must be independent of the network configuration chosen, and of the technology used to build the network, for a particular real system. A theory must apply equally well to Strowger switches, crossbar switches, gas-diode switches, and time-division switches. Unless it is independent of technology, a theory of connecting networks is limited in scope and may have missed the heart of the problem. We therefore use some of the terminology of switching engineers but understand it to refer to defined mathematical idealizations of switches, gates, crosspoints, transmission links, etc., rather than to the physical entities themselves.

We distinguish between switching networks used for communication and those used for control functions and logical transformations, like relay nets. Our concern is with networks of the former kind, and we call these *connecting networks.*

A communications switching network, or connecting network, consists of three kinds of entities: (*i*) *wires* or other transmission media along which communication may take place; (*ii*) *terminals* to which the wires are attached; and (*iii*) *crosspoints* or switches which can be used to connect the terminals, and hence the wires, together in various combinations. Each crosspoint can connect together exactly one pair of terminals, and it has two conditions: in the "on" or closed condition the two terminals are connected and communication can pass from one to the other; in the "off" or open condition the terminals are disconnected, and no information can pass through.

From the point of view of switching, two terminals connected together by a wire are essentially one terminal, albeit a spatially extended one. We therefore regard terminals as identical if they are wired together; in mathematical terms, we identify terminals under the equivalence relation of "being wired together." Henceforth, then, our considerations will leave wires out of account, and will be based only on the notions of *terminal* and *crosspoint.*

By the configuration or *structure* of a connecting network, we mean a specification of the terminals between which individual crosspoints have been placed. By the *condition* of a connecting network, we mean a specification of the closed and open crosspoints. In most cases of interest the structure is invariant in time, while the condition changes in a random way. We shall assume that at most one crosspoint is placed between distinct terminals, and that no crosspoint is placed from a terminal to itself.

IV. GRAPHICAL DEPICTION OF NETWORK STRUCTURE AND CONDITION

A simple device can illustrate the four notions we have introduced so far. In Fig. 1(a) the nodes (points) represent *terminals*, and the branches (lines) labeled x_i, $i = 1, \cdots, 6$, represent *crosspoints* placed between the terminals. The resulting graph represents the *structure* of a network. If we interpret the labels x_i as binary variables specifying the condition of the (respective) crosspoints, with 0 meaning "open" and 1 meaning "closed," then an assignment of values 0 or 1 to $\{x_1, \cdots, x_6\}$ represents a possible *condition* of the network, illustrated in Fig. 1(b). We are purposely avoiding the term "network state" here in order to assign it a useful precise meaning in the next section.

We have illustrated the use of a *labeled graph* as a general representation for (simultaneously) the structure and condition of a connecting

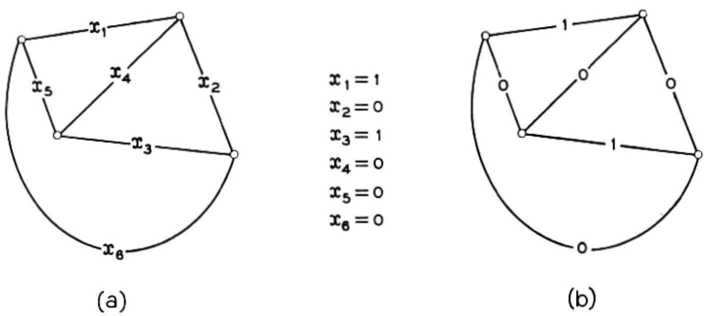

Fig. 1 — (a) Representation of structure; (b) simultaneous representation of structure and condition.

network. This representation is useful because it identifies the structure and condition of the network with a definite mathematical entity. It will become apparent that simple properties of this mathematical representation have great theoretical and practical relevance to congestion problems.

In general, the labeled graph g representing* the structure and condition of an arbitrary connecting network is constructed as follows:

i. nodes (points) of g correspond to terminals of the network;

ii. branches (lines, or edges) of g correspond to crosspoints of the network;

iii. open crosspoints are labeled 0;

iv. closed crosspoints are labeled 1.

Two terminals are *connected* in g if g contains a chain of closed crosspoints from one terminal to the other.

V. NETWORK STATES

Let G be a graph representing the structure of a switching network, and let V be the set of all labeled graphs g (labeled "versions" of G) obtained by assigning 0 or 1 to each line of G. There are several reasons why not every element g of V represents a physically meaningful state of the network.

In most switching systems there is an explicit functional distinction between terminals which are used only to connect other terminals together, and those between which desired connections arise, and which are never used to connect other terminals together. Terminals of the former kind we shall call *links,* because of their intermediary nature, and those of the latter kind, *inlets* and/or *outlets.* Desired connections always arise between two or more inlets or outlets. If more than two are involved, the connection is termed a "conference" call. Usually, though, the connections are disjoint chains of closed crosspoints, assuring private conversations between inlets and outlets by pairs only; we restrict attention to these. In terms of our graphical representation of the structure and condition of a switching network, the distinctions made above impose restrictions on the elements of V which represent realistic conditions of a network having the structure of G.

The restrictions on the assignment of the labels 0 or 1 listed above are (perhaps the most important) among many which are imposed by the functional and operational features of a real switching system. In general, a real connecting network specifies (or *uses*) only a subset of

* A glossary of mathematical notations appears at the end of this paper, Section IX.

the set V of all possible labeled versions of the graph G that represents the structure of the network being studied. We have therefore avoided calling elements of V "states of the network" because not all members of V can reasonably represent the condition of an actual network. We now attempt to characterize those subsets S of V which can represent real networks. Each such subset S will be called a *class of network states*.

The Boolean operations of *join* $\cup$ and *meet* $\cap$ (union and intersection, respectively) are definable for elements x, y of V in an obvious way:

$x \cup y =$ the V-element having a 1 wherever either x or y has a 1, and 0 elsewhere,

$x \cap y =$ the V-element having a 1 wherever both x and y have a 1, and 0 elsewhere.

The complement x' and the difference $x - y$ can be defined analogously. In view of this it is natural to inquire whether these Boolean operations can be used to characterize subsets S of V which are classes of network states.

If the elements x, y of V belong to such a subset (class of network states) S, it is not necessarily true that $x \cup y$, nor that $x \cap y$, belongs to S. In the case of $x \cup y$, there may be links and crosspoints used in both x and y, and so $x \cup y$ may violate the requirement of privacy. Even if $x \cap y = 0$ there may still be inlets used in both x and y, so that $x \cup y$ would lead to undesirable paths of extreme length. In the case of $x \cap y$, there may be so little in common to x and y that $x \cap y$ reduces to a single closed crosspoint between two links (i.e., *not* between an inlet and an outlet). Thus the Boolean operations do not yield a useful way of describing S.

The preceding remarks suggest that since any connection is a chain, none of whose terminals and crosspoints occurs in another connection, the labels 0 and 1 are really superfluous, although they served a tutorial purpose heretofore. That is, in describing the possible subsets S of network states, we can (and should) take advantage of inherent physical restrictions, and conveniently replace our representation* $x \in V$ of the structure and condition of a network by a corresponding set of disjoint chains, since each physically meaningful element x from V is equivalent to such a set. A formal development of this suggestion follows.

Let T be the set of terminals of a connecting network. The *graph* G representing the structure of the network is a subset G of the product

$$T \times T = \{(u, v) \mid u \in T, v \in T\}$$

* "$x \in V$" means that x is an element of the set V.

with the properties

$$(u, v) \in G \quad \text{if and only if} \quad (v, u) \in G$$

$$(u, u) \quad \text{is never in } G$$

and the interpretation

$(u, v) \in G$ if and only if (u, v) is an edge of graph G

if and only if nodes u and v are adjacent in the graph G

if and only if there is a crosspoint between terminals u and v.

A *chain* p of length n between terminals u and v is a sequence of elements $\{z_i \in T, 0 \leqq i \leqq n\}$ such that

$$z_0 = u, \qquad z_n = v,$$

$$z_i \neq z_j \quad \text{for} \quad i \neq j,$$

$$(z_i, z_{i+1}) \in G \quad \text{for} \quad i = 0, \cdots, n-1.$$

Two chains p_1 and p_2 are called *disjoint* if they have no nodes (terminals $\in T$) in common; in this case we write symbolically $p_1 \cap p_2 = \phi$, with ϕ = null set.

We shall henceforth assume that the set T of terminals has been (functionally) decomposed into three sets:

$$T = I \cup \Omega \cup L,$$

where I is a set of *inlets*, Ω a set of *outlets*, and L is the set of *links*. It is possible that $I = \Omega$ or that $I \cap \Omega$ = empty set, or that some intermediate condition obtain. However, we shall insist that $(I \cup \Omega) \cap L$ be null, i.e., that no link be an inlet or an outlet.

The set C of *connections* consists of all chains $p = \{z_i \in T, i = 0, \cdots, n(p)\}$ such that

$$z_0 \in I, \qquad z_{n(p)} \in \Omega, \qquad z_0 \neq z_{n(p)}$$

$$z_i \in L, \quad \text{for} \quad i \neq 0 \quad \text{or} \quad n(p).$$

Each element p of C represents a possible connection from an inlet to an outlet through the network whose structure is represented by the graph G.

Elements of the set S of network states will be defined as subsets x of C, $x \subset C$, consisting entirely of disjoint chains, that is, such that

$$p_1, p_2 \in x \quad \text{implies} \quad p_1 \cap p_2 = \phi.$$

Two subsets x and y of C are called *compatible* if

$$p_1 \in x,\ p_2 \in y \quad \text{implies} \quad p_1 \cap p_2 = \phi.$$

The connections that comprise compatible states can all be put up simultaneously without interfering with each other or violating the requirement of privacy.

The functional and physical restrictions imposed by real networks determine (in any particular system) a subset E of C consisting of (what we shall call) the *elementary states*, or single connections that can actually be used. For example, chains in C that double back and are wastefully circuitous may be excluded from E.

Given such a subset E of elementary states, we can define a class of network states S, associated with E, in a natural way as follows: S is the smallest class of subsets of E containing all unit subsets of E, and closed under formation of arbitrary intersections (meets) and unions (joins) of *compatible* subsets of E. That is, S is the smallest class of E-subsets such that

$$p \in E \quad \text{implies} \quad \{p\} \in S,$$

$$x, y \in S \quad \text{implies} \quad x \cap y \in S,$$

$$\text{if} \quad x, y \in S \quad \text{and} \quad p_1 \in x,\ p_2 \in y \quad \text{implies} \quad p_1 \cap p_2 = \phi,$$
$$\text{then} \quad x \cup y \in S.$$

We henceforth use "S" as a generic notation for a class of network states defined as above. The word "network" will refer to a graph G representing structure, choices I and Ω of inlets and outlets respectively, and a choice E of elementary states. The choice of G, I, Ω, and E uniquely determines a class S of network states according to the definition given previously. The quadruple (G, I, Ω, E) will be called a network, N.

It is easily verified that the class S of network states is partially ordered by inclusion, $\leqq$. Moreover, any two elements x, y of S have a unique intersection (meet) consisting of just those connections common to both x and y, and S itself has a unique least element included in every other element, viz., the ground state in which no calls are in progress. However, since only infima exist, and since there may be many maximal elements in the partial ordering, S is not a lattice, in general.

VI. THE STATE DIAGRAM

The partial ordering $\leqq$ of S has a special nature that allows us to arrange the network states $x \in S$ in a particularly intuitive and useful pattern. The following conventions and definitions will be helpful in discussing this pattern.

If K is any set, we use the notation $|K|$ to mean the number of elements of K. E.g., if x is a network state,

$$|x| = \text{the number of calls in progress in state } x.$$

The sets L_k are defined by the conditions

$$L_k = \{x \in S \mid |x| = k\}, \qquad k = 0, 1, \cdots,$$

that is, L_k is the set of all network states consisting of exactly k connections. L_0 is a unit set containing just the zero state. The sets L_k are a partition of S corresponding to the equivalence relation of "having the same number of calls in progress."

To obtain our pattern for arranging network states we start with the zero or ground state in which no calls are in progress: this is the empty set (of chains). Above this zero element, in a horizontal row, we place all the states consisting of a single connection, i.e., all the elements of E. Continuing in this way, we put the set L_{k+1} of states consisting of $(k + 1)$ disjoint chains (i.e., $k + 1$ calls) in a horizontal row above the set L_k of states with k disjoint chains (i.e., k calls in progress). We call L_k the kth *level*.

The diagram is completed by constructing the corresponding Hasse figure (Birkhoff,[3] p. 12); that is, we think of the states $x \in S$ (in their arrangement into levels L_k) as nodes, and we construct a graph by drawing lines between states x, y of respective adjacent levels L_k, L_{k+1} just in case

$$y - x \in E,$$

i.e., if and only if y results from x by putting up one more call. The resulting graph can be termed the *state diagram* D of the network N described by the quadruple (G, I, Ω, E). The state diagram D is a natural and standard representation of the partial ordering of S. The history of the connecting network when operating can be thought of as a trajectory on D.

We shall use the network depicted in Fig. 2 to illustrate the state diagram D. For most practical purposes this network is wasteful of crosspoints, but it makes a suitably simple example of the partial ordering of the states. The network has four inlets and four outlets, and no inlet is an outlet. The squares in Fig. 2 represent 2 by 2 switches, as indicated.

The possible states of this network are determined by all the ways

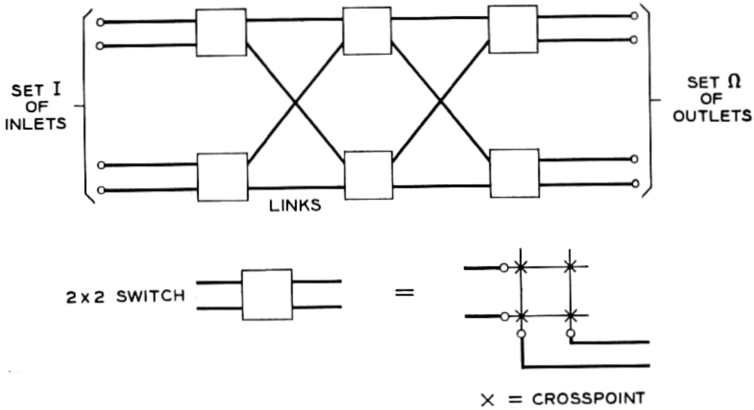

Fig. 2 — Illustrative three-stage connecting network.

in which four or fewer inlets can be connected pairwise to as many outlets on the right, no inlet being connected to more than one outlet, and vice versa. These possible states have been depicted in a natural arrangement in Fig. 3, which shows a reduced state diagram in which states which differ only by permutations of inlets, outlets, or switches have been identified. There is essentially only one way to put in a single call; there are four ways of putting in two calls; and there are two ways each of putting in three and four calls. The states have been arranged in levels according to the number of calls in progress. In each state only links actually in use are shown, and the different notations on the links indicate the routing.

VII. SOME NUMERICAL FUNCTIONS

The finite set S of network states is *partially ordered* by *inclusion*, which we shall denote by $\leqq$. A *chain* in S is a subset X of S which is *simply ordered* by (the restriction to X of) $\leqq$; that is, for any two elements $x, y \in X$, we have either $x \geqq y$ or $y \geqq x$. Such a chain is not to be confused with the "chains" on the graph G that are elements of states $x \in S$. The *dimension* or *height* $|x|$ of a state is the maximum "length" d of chains $0 < x_1 < \cdots < x_d = x$ that have x for greatest element. (This usage is consistent with the previous definition of $|\cdot|$.)

Remark 1: The dimension $|x|$ of a state x is the *number of busy pairs*, or the *number of calls in progress*, in the state x.

A state x is said to *cover* another state y if and only if $x > y$, and there are no $z \in S$ such that $x > z > y$. The state x is then "immediately above" y. It is apparent that x covers y if and only if $x > y$ and $|x| =$

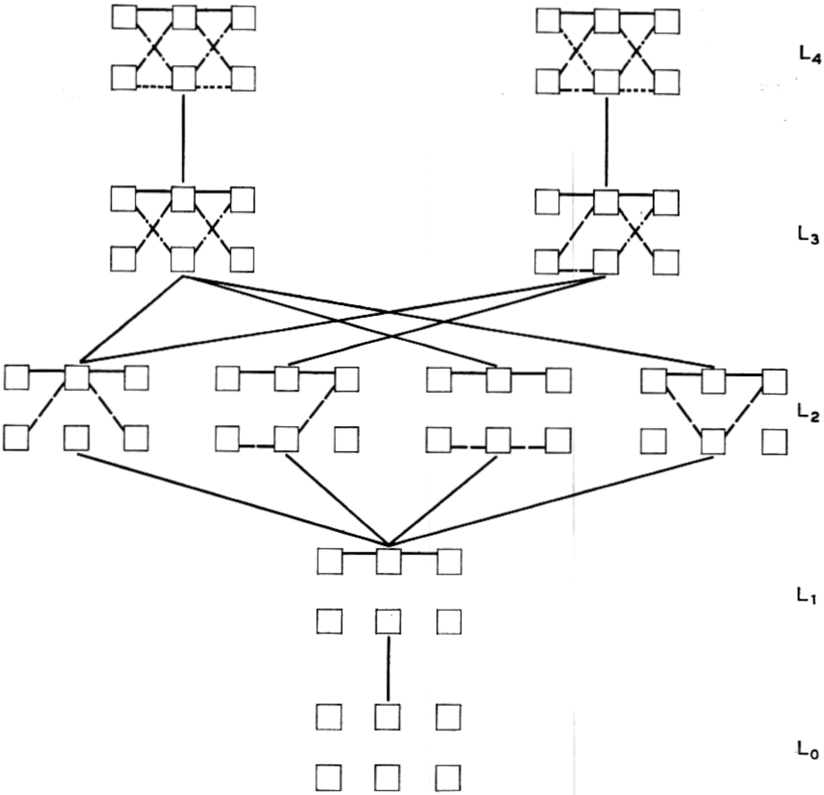

Fig. 3 — (Reduced) state diagram for the network shown in Fig. 2.

$|y| + 1$. In fact, the construction of the partial ordering of S arranges the states according to levels, each level being the (equivalence) class of all states having the same *dimension*. In determining dimension one need only consider chains that are "maximal" or "connected" in the sense that x_i covers x_{i-1} for all i. Also, it can be seen that the partial ordering $\leqq$ of S satisfies the *Jordan-Dedekind chain condition*: all connected chains between fixed end points have the same length.

The present section will be devoted to various relationships between numerical functions defined on S, counting or "enumeration" problems, etc., based largely on the dimension function and the chain condition.

The *Möbius function* $\mu(\cdot)$ of the partially ordered system $(S, \leqq)$ is defined recursively by

$$\mu(0) = 1, \qquad \mu(x) = -\sum_{y<x} \mu(y) \quad \text{if} \quad x > 0,$$

where 0 denotes the zero or ground state in which no calls are up. The Möbius function has the following two important properties:

i. Let $f(\cdot)$ be any function defined on S, and let

$$F(x) = \sum_{y<x} f(y).$$

Then $f(\cdot)$ and $F(\cdot)$ are related by the Möbius inversion formula

$$f(x) = \sum_{y<x} \mu(y)F(x - y).$$

Here $x - y$ denotes the state obtained from x by removing all the calls of state y; this makes sense, since $y < x$. (See Weisner.[4])

ii. Let $\lambda(x, n)$ be the number of chains of length n which can be interpolated between 0 and x. Then P. Hall[5] has shown that

$$-\mu(x) = \lambda(x, 1) - \lambda(x, 2) + \cdots .$$

By the Jordan-Dedekind chain condition, all the chains from 0 to x have the same length, viz., $|x|$. Hence for $x > 0$

$$\mu(x) = (-1)^{|x|} \lambda(x, |x|).$$

For simplicity of notation set

$$\begin{aligned} \lambda(x, |x|) &= \eta(x) \\ &= \text{number of ways of ``climbing'' from 0 to } x. \end{aligned}$$

Also, we introduce the following sets:

$$A_x = \{y \mid y \text{ covers } x\}$$

$$B_x = \{y \mid x \text{ covers } y\}$$

$$L_n = \{x \mid |x| = n\}.$$

These have the following respective intuitive meanings: A_x is the set of states *immediately above* x, i.e., obtainable from x by adding one more call; B_x is the set of states immediately below x, i.e., obtainable from x by removing one call; L_n is the nth level, the set of all states having n calls up. The cardinality of a finite set X is designated by $|X|$.

Remark 2: $|B_x| = |x|$ for each $x \in S$. Clearly, x covers exactly $|x|$ states, each obtainable from x by removing one call.

Remark 3: For each $x \in S$

$$\eta(x) = \sum_{y \in B_x} \eta(y).$$

Indeed, every state y covered by x gives rise to exactly $\eta(y)$ climbing paths from 0 that reach x via y.

Remark 4: For $x \epsilon L_n$, $\eta(x)$ has the constant value $n!$ This is obvious intuitively, since there are $n!$ orders in which the n calls of $x \epsilon L_n$ could be put up. More formally, the result is true for $x = 0$; assume it true for $y \epsilon L_{n-1}$; then by the previous results,

$$\eta(x) = \sum_{y \epsilon B_x} \eta(y) = | B_x | \cdot (n - 1)!$$
$$= n!.$$

Remark 5: The Möbius function $\mu(\cdot)$ is given by

$$\mu(x) = (-1)^{|x|} (| x |)!$$
$$= (-1)^n n! \quad \text{for} \quad x \epsilon L_n .$$

Theorem 1:

$$| L_n | = \frac{1}{n} \sum_{y \epsilon L_{n-1}} | A_y |, \qquad n > 0.$$

Proof: The segments in the partial ordering passing from elements $y \epsilon L_{n-1}$ to L_n are just those that pass from some $x \epsilon L_n$ to L_{n-1} and by Remark 2, each $x \epsilon L_n$ has exactly $| x |$ $(= n)$ such segments. Therefore,

$$n \cdot | L_n | = \sum_{y \epsilon L_{n-1}} | A_y |$$

and the sum on the right is exactly divisible by n.

Definition: C_n is the total number of chains (of length n) from 0 into L_n, i.e., to some state in L_n.

Remark 6:

$$C(n) = \sum_{x \epsilon L_n} \eta(x) = \sum_{y \epsilon L_{n-1}} \eta(y) \cdot | A_y |.$$

It can be seen that $x \epsilon L_n$ has $\eta(x)$ chains climbing to it from 0; for $x, y \epsilon L_n$, $x \neq y$, these chains are distinct since their highest elements are unequal. This proves the first identity. Also each chain climbing to L_n from 0 must pass through some unique $y \epsilon L_{n-1}$. Each $y \epsilon L_{n-1}$ has $\eta(y)$ chains of length $n - 1$ reaching it from 0, and each such chain can then be completed to reach L_n in $| A_y |$ ways. It follows also that

$$| L_n | = \frac{C_n}{n!} = \frac{1}{n} \sum_{y \epsilon L_{n-1}} | A_y | .$$

VIII. ASSIGNMENTS

By an *assignment* we shall mean any one-to-one map $a(\cdot)$ of a subset of I into Ω. An assignment is to be interpreted as a specification of what inlets are to be connected to what outlets, without regard to the possible routes that these connections might take through the network. If $I \cap \Omega$ is nonnull, we restrict assignments so as to satisfy $a(u) \neq u$.

Let x be a network state consisting of chains $p_1, p_2, \cdots, p_n$ with $n = n(x)$ and each p_i a chain between $u_i \in I$ and $v_i \in \Omega$. We say that *x realizes the assignment* $a(\cdot)$ if and only if

i. the domain of $a(\cdot)$ is $\{u_i, 0 \leqq i \leqq n(x)\}$

ii. the range of $a(\cdot)$ is $\{v_i, 0 \leqq i \leqq n(x)\}$

iii. $a(u_i) = v_i, 0 \leqq i \leqq n(x)$.

An assignment is *realizable* if some network state realizes it; a state realizes exactly one assignment; the zero state realizes the null assignment. A *maximal* assignment is one that has either domain I or range Ω. The set of all assignments is denoted by A, and that of all maximal assignments by $\hat{A}$.

Two terminals, $u \in I$ and $v \in \Omega$, are *connected* in state x if and only if some chain $p \in x$ is a chain between u and v, i.e., if and only if x realizes the (unit) assignment

$$\{(u, v)\}.$$

We define the function $\gamma(\cdot)$ from S into (the set of) subsets of $I \times \Omega$ by the condition

$$\gamma(x) = \{(u, v) \in I \times \Omega \mid u \text{ and } v \text{ are connected in } x\}.$$

Formally, then $\gamma(x)$ is the assignment realized by state x; heuristically, we may think of $\gamma(x)$ as the set of calls which are in progress in state x. The set of *unit assignments*, that is, of

$$c = \{(u, v)\} \quad \text{such that} \quad (u, v) \in I \times \Omega,$$

will be denoted by U, and a unit assignment $c \in U$ will be referred to informally as a *call*.

If $a = a(\cdot) \in A$ is an assignment, we use the notation

$$\gamma^{-1}(a)$$

for the inverse image of $a(\cdot)$ under $\gamma(\cdot)$, i.e., the set of (equivalent) states y such that $\gamma(y) = a$. In a similar vein, if X is a set of states, we define

$$\gamma(X) = \{a \in A \mid a = \gamma(x) \quad \text{for some} \quad x \in X\},$$

that is, $\gamma(X)$ is the set of assignments realized by members of X.

IX. THREE TOPOLOGIES

Two network states x and y are equivalent, written $x \sim y$, if and only if they realize the same assignment, i.e.,

$$\gamma(x) = \gamma(y).$$

Intuitively, equivalent but nonidentical states correspond to different ways of putting up the same set of calls.

A pseudo-metric (Kelley,[6] p. 118) on S can be defined by the formula

$$d(x, y) = |\gamma(x)\Delta\gamma(y)|, \qquad x, y \in S,$$

where Δ denotes the symmetric difference of sets, and $|\cdot|$ cardinality, as before. In plain words, the distance $d(x, y)$ between x and y is the number of pairs $(u, v) \in I \times \Omega$ that are either connected in x and not connected in y, or connected in y and not connected in x. Clearly

$$d(x, 0) = |x|, \qquad 0 = \text{zero state},$$

and also

$$d(x, y) = 0 \quad \text{if and only if} \quad x \sim y.$$

Thus $d(\cdot, \cdot)$ only identifies states up to equivalence. The function $d(\cdot, \cdot)$ is obviously symmetric, and the triangle inequality is a consequence of the set inclusion

$$(X\Delta Y) \subseteq (X\Delta Z) \cup (Y\Delta Z).$$

The pseudo-metric $d(\cdot, \cdot)$ can be used to define a topology for S in a standard way (see Kelley,[6] p. 118 et seq.) The *closure* of a set X in the d-topology consists of all states equivalent to members of X, and is denoted by X^d.

For each subset X of S, we define its $\leqq$-*closure* $\underline{X}$ by the condition

$$\underline{X} = \{y \in S \mid y \leqq x \quad \text{for some} \quad x \in X\}.$$

The operation on sets so defined satisfies the Kuratowski closure axioms (cf. Kelley,[6] p. 43):

$$\underline{\phi} = \phi$$

$$X \subseteq \underline{X}$$

$$\underline{\underline{X}} = \underline{X}$$

$$\underline{X \cup Y} = \underline{X} \cup \underline{Y}$$

and so defines a closure topology for S. The set $\underline{X}$ consists of all states that are "below" some member of X in the state-diagram D, i.e., can be reached from a member of X by removing calls.

In a similar way, we define the $\geqq$*-closure* $\bar{X}$ of a set $X \subseteq S$ as

$$\bar{X} = \{y \in S \mid y \geqq x \quad \text{for some} \quad x \in X\}.$$

The converse of a partial ordering relation is also a partial ordering, called its *dual*. Hence the mapping $X \to \bar{X}$ is also a closure operation, defining a third topology on S.

X. SOME DEFINITIONS AND PROBLEMS

An inlet or outlet is *idle* in a network state x if it belongs to neither the range nor the domain of the assignment $\gamma(x)$ realized by x. An *idle pair* of the state x is an element (u, v) of $I \times \Omega$ such that both u and v are idle in x. A call $c = \{(u, v)\}$ is *new* in x if (u, v) is an idle pair.

We shall now define what is meant by a blocked call. Let $x \in S$ realize the assignment $\gamma(x)$ and let c be a new call in x, i.e., let

$$c = \{(u, v)\} \in U$$

be a unit assignment such that (u, v) is an idle pair of x. The new call c is *blocked in* x if there is no state $y > x$ such that

$$\gamma(y) = \gamma(x) \cup c.$$

A state x is a *blocking state* if some call is blocked in x. The state x is called *nonblocking* if and only if for every idle pair (u, v) of x, the call

$$c = \{(u, v)\}$$

is not blocked in x, i.e., there is a $y \in S$ above x which realizes the larger assignment $\gamma(x) \cup c$, so that

$$\gamma(y) = \gamma(x) \cup \{(u, v)\},$$
$$y > x.$$

The set of nonblocking states is designated by the symbol B'. A state that realizes a maximal assignment has no idle pairs, and is (trivially) nonblocking. In plain terms, a nonblocking state x is one in which any idle inlet u can be connected to any idle outlet v *without disturbing the calls that are already present*; in this case there is a path r, disjoint from all paths $p \in x$, between u and v, and

$$x \cup \{r\} \in S,$$

i.e., use of this path results in a network state.

A network $N = (G, I, \Omega, E)$ will be called *nonblocking in the strict sense* if and only if every state is nonblocking, i.e., $B' = S$. Such networks have been discovered and studied extensively by C. Clos. (See Clos[7] and Kharkevich.[8]) A network that is nonblocking in this strong sense has the property that no matter in what state it is, any idle pair can be connected (in a way that results in a legitimate network state).

In most switching networks there may be several or many ways of connecting an idle pair, i.e., putting up a new call, in a given state, all of which lead to legitimate network states. Thus, even if the set S of network states contains blocking states, it is conceivable that by making the right choices of paths for connections one might avoid all the blocking states, and still satisfy all demands for connection as they arise, without disturbing calls already present. That is, there may exist a *rule* for choosing paths which, if followed, confines the trajectory of the system to nonblocking states (without refusing any demands for connection by idle pairs).

We next discuss what is meant by a rule. If a call $c = \{(u, v)\}$ is blocked in a state x it cannot be put up without disturbing existing calls of x, and there is no question of using a rule. Also, if x is a maximal state, no new calls can be put up, and a rule is unnecessary. But if a call c can be put up in one or more ways in the state x, then there is at least one $y > x$ such that $\gamma(y) = \gamma(x) \cup c$. In such a case some method of specifying permitted or prohibited new states could be used in order to improve performance.

A *rule* $\rho(\cdot\ ,\ \cdot)$ for a network N is a mapping of the Cartesian product

$$[S - \gamma^{-1}(\hat{A})] \times U$$

into subsets of S, with the properties: if $x \in S$ and $c = \{(u, v)\} \in U$ with (u, v) an idle pair of x (so that c is a new call in x), then

$$0 \subseteq \rho(x, c) \subseteq \gamma^{-1}(\gamma(x) \cup c);$$

if x is maximal, or if (u, v) is not idle, $\rho(x, c)$ is defined (arbitrarily) as the null set. If for some call c not up in x we have

$$y \in \rho(x, c),$$

we say that the transition (between states) $x \rightarrow y$ is *permitted* by $\rho(\cdot\ ,\ \cdot)$.

We say informally that a state x is *reachable under a rule* $\rho(\cdot\ ,\ \cdot)$ if there is some sequence of changes of state, consisting of either hangups or transitions permitted by $\rho(\cdot\ ,\ \cdot)$, and leading from the zero state to x. More precisely, we define the notion

x is reachable under $\rho(\cdot\ ,\ \cdot)$ in n steps

recursively, as follows:

i. The zero state is reachable under $\rho(\cdot\ ,\ \cdot)$ in zero steps.

ii. If x is reachable under $\rho(\cdot\ ,\ \cdot)$ in n steps, and for some call $c \in U$, $\gamma(x) = \gamma(y) \cup c$, then y is reachable under $\rho(\cdot\ ,\ \cdot)$ in $(n + 1)$ steps.

iii. If x is reachable under $\rho(\cdot\ ,\ \cdot)$ in n steps, and for some call $c \in U$, c is new in x and $y \in \rho(x, c)$, then y is reachable under $\rho(\cdot\ ,\ \cdot)$ in $(n + 1)$ steps.

A state is *reachable under* $\rho(\cdot\ ,\ \cdot)$ if it is reachable under $\rho(\cdot\ ,\ \cdot)$ in n steps, for some $n \geqq 0$. The set of states that are reachable under $\rho(\cdot\ ,\ \cdot)$ will be denoted by R_ρ.

A network $N = (G, I, \Omega, E)$ will be called *nonblocking in the wide sense* if and only if there is a rule $\rho(\cdot\ ,\ \cdot)$ for N under which no blocking state is reachable, i.e.,

$$R_\rho \subseteq B'.$$

In words, we may say that a network is nonblocking in the wide sense if there is a rule, depending on the states, and on the connections that are requested, such that if the rule if used (starting from the zero state) no blocking state is ever reached, and hence no request for connection by an idle pair (of a state that can be reached) need ever be refused. In making this definition, we think of the system as starting (empty) at the zero state; in any state x that it reaches, any idle pair of x may demand connection; it must always be possible to make this connection without disturbing existing calls, and reach a (nonblocking) state y one level higher, $y \in L_{|x|+1}$; at any instant an existing call may terminate, and the system move to a state of $L_{|x|-1}$. An example of such a network was given in Ref. 1.

Finally, we consider a still weaker property of networks than the first two defined, namely, the possibility of satisfying a demand for connection by *rearranging* (if necessary) the existing calls in such a way that the desired call can then be accommodated. Let x be a network state realizing the assignment $\gamma(x)$. We call x *rearrangeable* if and only if for every idle pair (u, v) of x there is a $y \in S$, possibly depending on (u, v) and x, which realizes the larger assignment $\gamma(x) \cup \{(u, v)\}$, i.e.,

$$\gamma(y) = \gamma(x) \cup \{(u, v)\}.$$

Alternately x is rearrangeable if for every call c new in x there is a y such that

$$\gamma(y) = \gamma(x) \cup c.$$

This definition is the same as that of a nonblocking state except that the condition $x < y$ is omitted. That is, to realize the larger assignment $\gamma(x) \cup c$ it may be necessary to reroute existing calls to give a new state $z \sim x$ which is not comparable to x, and which has a path r, disjoint from $p \in z$, between u and v. The state y may then be taken to be $z \cup \{r\}$. A network N is called *rearrangeable* if its states $x \in S$ are rearrangeable.

With these definitions laid down, we can formulate several problems of the combinatory theory of connecting networks:

i. Can *general* characterizations of the properties of being rearrangeable, and of being nonblocking (strict or wide sense) be given?

ii. What relationships exist among the concepts we have defined?

iii. What *specific* networks are rearrangeable, or non-blocking (strict or wide sense)?

To attack problem (*i*) we make the following observations: the three properties of interest represent different degrees of abundance of states in which calls are not blocked. The relative abundance or density of such states throughout S determines which (if any) of the three properties N has. The heuristic concept of abundance suggests the topological one of *denseness*, and the possibility of characterizing the three properties in terms of denseness. This idea is developed in the remaining sections; it leads to answers to problems (*i*) and (*ii*) above.

XI. REARRANGEABLE NETWORKS

Let X be a subclass of the class S of network states. We say that X is *sufficient* if $\gamma(X) = A$, i.e., if every assignment is realized by some state of X. We make two comments:

Remark 7: If $\hat{A} \subseteq \gamma(x)$, then $\underline{X}$ is sufficient. This can be seen as follows: every assignment is a subset of some maximal assignment, and so belongs to the $\leqq$-closure $\underline{X}$ of X. For the same reason we have

Remark 8: The following properties of a network N are equivalent:

i. N is rearrangeable.

ii. Some sufficient class exists.

iii. The range of $\gamma(\cdot)$ includes $\hat{A}$.

It is convenient to approach the study of rearrangeable networks by taking the point of view of a particular pair of customers, i.e., of a particular inlet-outlet pair $(u, v) \in I \times \Omega$. Such a pair corresponds to a unit assignment or *call*

$$c = \{(u, v)\} \in U,$$

any realization of which is among the states of E, the set of elementary states. For each call $c \in U$ we define

$$I_c = \{x \in S \mid c \text{ is new in } x, \text{ i.e., } (u, v) \text{ is idle in } x\},$$

$$B_c = \{x \in S \mid c \text{ is blocked in } x\}.$$

It can be verified that

$$B_c \subset I_c, \qquad \text{for} \quad c \in U,$$

$$B' = \bigcap_{c \in U} (B_c)',$$

$$S - \gamma^{-1}(\hat{A}) = \bigcup_{c \in U} I_c .$$

We call a network N rearrangeable for the unit assignment or call c if and only if for every $x \in I_c$ there is a $y \in S - I_c$ which realizes the larger assignment $\gamma(x) \cup c = \gamma(y)$. In words, this condition states that for any state in which the pair (u, v) is idle there is a (possibly rearranged) state in which all the same calls are up, and in addition u is connected to v. It is easy to see that N is rearrangeable if and only if it is rearrangeable for all calls $c \in U$.

Let X, Y be arbitrary subsets of S. In accord with a standard definition (Kelley,[6] p. 49), X is said to be dense in Y in the d-metric if Y is included in the d-closure of X, i.e.,

$$Y \subseteq X^d.$$

Now in a metric space the closure of a set X is the set of all points that are at distance zero from X, when the distance of a point y from a set X is defined as

$$\inf_{x \in X} d(x, y) .$$

Hence the closure of X is the set of all y such that for some $x \in X$, $d(x, y) = 0$, or equivalently, $x \sim y$. That is, the d-closure of X is the set of all states that are equivalent to a member of X:

$$X^d = \{y \in S \mid y \sim x \quad \text{for some} \quad x \in X\}.$$

These observations lead to the following result:

Theorem 2: N is rearrangeable if and only if

$$(B_c)' \text{ is } d\text{-dense in } I_c , \qquad \text{for each } c \in U.$$

Proof: Let N be rearrangeable; let $c \in U$; and pick x in I_c. Then there exists $y \in S$ such that

$$\gamma(y) = \gamma(x) \cup c,$$

and so there exists a $z \in E \cap \gamma^{-1}(c)$ such that $z \leqq y$ and $x \sim y - z$. Obviously then

$$y - z \in (B_c)'$$

and since x is equivalent to $y - z$ we have

$$x \in ((B_c)')^d.$$

Since x was an arbitrary member of I_c, we have proved $I_c \subseteq ((B_c)')^d$. Conversely, assume that the condition in the theorem holds, and pick any $c \in U$, and $x \in I_c$. Then $x \sim y$ for some y in $(B_c)'$, so that c is not blocked in y. Thus N is rearrangeable for all $c \in U$, and so is rearrangeable.

A similar argument yields the weaker and simpler result:

Remark 9: If B' is d-dense in S, then N is rearrangeable. In this case, since $S \subseteq (B')^d$, given a state x there is always an equivalent nonblocking state y, with

$$y \sim x,\ y \in B'.$$

Hence rearrangements can be made uniformly in the calls new to x.

XII. NETWORKS NONBLOCKING IN THE WIDE SENSE

We now turn to the characterization of networks for which there is a rule for routing calls which allows the operating system to avoid blocking states entirely. The case in which the network is actually nonblocking in the strict sense, so that *any* rule will do, is excluded here as trivial. The point is to use a network with blocking states, but to manage to avoid them by clever routing. The following general criterion of a useful rule $\rho(\cdot\ ,\ \cdot)$ suggests itself: $\rho(\cdot\ ,\ \cdot)$ should make as many blocking states as possible unreachable, consistent with satisfying requests for connection by unblocked new calls.

To exhibit, in an intuitive way, all the relationships that obtain, it is convenient to introduce an additional concept: a class X of network states is *preservable* (*by new calls*) if and only if for any $x \in X$ and any call c that is new to x and unblocked in x, there is a state $y \in X$ such that

$$y > x \quad \text{and} \quad \gamma(y) = \gamma(x) \cup c.$$

That is, if an idle pair (u, v) of x corresponds to a call $c = \{(u, v)\}$ that is unblocked in x, then some state $y \in X$ realizes $\gamma(x) \cup \{(u, v)\}$, and y is above x in the state-diagram, $y > x$. In words, X is preservable if any call that can be put up at all in a state of X can be put up *salva* staying in X, that is, in such a way that the system stays in X. A $\geqq$-closed class is always preservable (by new calls). We make

Remark 10: If X is preservable, $0 \epsilon X$, and $X \subseteq B'$, then X is sufficient.

It is then possible to start at the zero state, and call by call realize any maximal assignment *salva* staying in X. We now state

Theorem 3: N is nonblocking in the wide sense if and only if there exists a nonempty subset X of states such that

i. X is preservable.

ii. $X \subseteq B'$.

iii. X is $\leqq$-closed, i.e., $X = \underline{X}$.

Proof: Let (*i*)–(*iii*) hold for some subset X, and define a rule $\rho(\cdot\ ,\ \cdot)$ by the condition that if $c \epsilon U$ is new to x, then

$$\rho(x, c) = \gamma^{-1} (\gamma(x) \cup c) \cap X.$$

Use of $\rho(\cdot\ ,\ \cdot)$ is tantamount to requiring that any call must be put up so as to lead to a state of X. By (*i*) and (*ii*), this can always be done. Since X is $\leqq$-closed, hangups preserve membership in X; since X is nonempty it contains the zero state. Hence all states reachable under $\rho(\cdot\ ,\ \cdot)$ belong to $X \cap B'$ and

$$R_\rho \subseteq B',$$

so that N is nonblocking in the wide sense.

Conversely, if N is nonblocking in the wide sense, then some rule $\rho(\cdot\ ,\ \cdot)$ is such that no blocking state belongs to R_ρ. Set $X = R_\rho$. Then X is $\leqq$-closed, because any state below a reachable state is reachable by hangups. Also $X \subseteq B'$, because $\rho(\cdot\ ,\ \cdot)$ avoids all blocking states. Finally, X must be preservable since one can "preserve" X simply by using only state-transitions permitted by $\rho(\cdot\ ,\ \cdot)$, i.e., by putting up unblocked new calls so as to lead only to states vouchsafed by $\rho(\cdot\ ,\ \cdot)$.

We recall that for $x \epsilon S$,

$$\begin{aligned} A_x &= \{y \mid y \text{ covers } x\}, \\ &= \{y \mid y = x \cup z \text{ for some } z \epsilon E\}, \\ &= \{\text{set of states immediately above } x\}. \end{aligned}$$

The property of preservability (of a set X of states) will now be given a topological characterization in terms of *denseness*, in the following result:

Theorem 4: A nonempty subset X of S is preservable if and only if for every $x \epsilon X$, $A_x \cap X$ is dense in A_x in the sense of the d-metric; i.e., $x \epsilon X$ implies

$$A_x \subseteq (A_x \cap X)^d.$$

Proof: Take $x \epsilon X$ and $y \epsilon A_x$, so that y is "immediately above" x, or y covers x. Then there is a call c new in x such that

$$\gamma(y) = \gamma(x) \cup c,$$

and so if $A_x \cap X$ is dense in A_x, there is a $z \epsilon A_x \cap X$ which is equivalent to y. Since z covers x, it follows that the call c new to x can be connected in state x so as to give rise to a state of X. That is, we have

$$z \epsilon X$$

$$\gamma(z) = \gamma(x) \cup c.$$

Since c was an arbitrary new call of $x \epsilon X$, the set X is preservable, if the condition of Theorem 4 is true. Conversely, let X be preservable, and take $x \epsilon X$ and $y \epsilon A_x$. Then there exists a call c not blocked in x with $\gamma(y) = \gamma(x) \cup c$. But since X is preservable, and c is not blocked in x, there is a z in $A_x \cap X$ such that $\gamma(z) = \gamma(x) \cup c$, that is $z \sim y$. Hence y is equivalent to an element of $A_x \cap X$. Since y was arbitrary, it follows that for $x \epsilon X$,

$$A_x \subseteq (A_x \cap X)^d.$$

Remark 11: The sets $\{A_x, x \epsilon X\}$ in the condition of Theorem 4 may be replaced by the "x-cones"

$$\{y \mid y > x\}.$$

XIII. NETWORKS NONBLOCKING IN THE STRICT SENSE

A network that is nonblocking in the strict sense has no blocking states whatever. A simple characterization of this property is given by

Theorem 5: N is nonblocking in the strict sense if and only if there is a subset X of B' such that

i. X is sufficient

ii. X is d-closed, i.e., $X = X^d$.

Proof: If N has the property, then $S = B'$, and we may take $X = S$. Conversely, if (*i*) and (*ii*) obtain, take any $x \epsilon S$; since X is sufficient, there exists $y \epsilon X$ for which $\gamma(x) = \gamma(y)$, i.e., $x \sim y$. But $X = X^d$, so $x \epsilon X$, and hence $x \epsilon B'$.

By Remark 10, the condition (i) that X be sufficient can be replaced by the condition that X be preservable and nonempty.

IX. GLOSSARY

G	an arbitrary graph
g	a copy of G with each edge labeled 0 or 1
V	the set of all labeled versions g of G
S	the set of network states (typical members x, y, z)
T	the set of terminals (nodes of G)
p	a typical path (chain) on G
ϕ	null set
I	the set of inlets
Ω	the set of outlets
L	the set of links
C	the set of all connections (paths from I to Ω)
E	the set of elementary states
N	an arbitrary network, specified by choosing G, I, Ω, and E
$\leqq$	partial ordering of V or S by inclusion
0	zero state
$\vert x \vert$	number of elements of the set X
L_k	set of states with exactly k calls up
D	state diagram (Hasse figure of $\leqq$ on S)
$\mu(\cdot)$	Möbius function of $\leqq$
$\lambda(x, n)$	number of chains of length n from 0 to x
$\eta(x)$	$\lambda(x, \vert x \vert)$
A_x	set of states directly above x
B_x	set of states directly below x
C_n	$\sum_{x \in L_n} \eta(x)$
$a(\cdot)$	an assignment (any 1–1 map of subset of I into Ω)
$\hat{A}$	set of maximal assignments
U	set of unit assignments or calls
c	a call, or typical member of U
$\gamma(x)$	the assignment realized by state x
$\sim$	equivalence of states
$d(x, y)$	$\vert \gamma(x) \Delta \gamma(y) \vert$
$\underline{X}$	$\leqq$-closure of X
$\bar{X}$	$\geqq$-closure of X
X^d	d-closure of X
B	set of states in which some call is blocked
$\rho(\cdot, \cdot)$	a rule for operating a network
R_ρ	the set of states reachable under $\rho(\cdot, \cdot)$.

REFERENCES

1. Beneš, V. E., Heuristic Remarks and Mathematical Problems Regarding the Theory of Switching Systems, this issue, pp. 1201–1247.
2. Beneš, V. E., On Rearrangeable Three-Stage Connecting Networks, to appear.
3. Birkhoff, G., Lattice Theory, Amer. Math. Soc. Colloq. Publ. XXV, rev. ed., 1948.
4. Weisner, L., Trans. Amer. Math. Soc., **38,** 1955, pp. 474–484.
5. Hall, P., Quarterly Journal of Mathematics, **7,** 1936, pp. 134–151.
6. Kelley, J. L., *General Topology*, D. Van Nostrand, New York, 1955.
7. Clos, C., A Study of Non-Blocking Switching Networks, B.S.T.J., **32,** 1953, pp. 406–424.
8. Kharkevich, A. D., Multi-Stage Construction of Switching Systems (in Russian), Doklady AN SSSR, **112,** 1957, pp. 1043–6.