# On Some Proposed Models for Traffic in Connecting Networks

By V. E. BENEŠ

(Manuscript received August 17, 1966)

*Three stochastic models for traffic, forming a progression of decreasing simplicity, are discussed with a view to discerning in what ways the various assumptions they depend on affect the formula for blocking probability. These models are the probability linear graph (due to C. Y. Lee), the thermodynamic model, and a model based on Markov processes (both proposed by the author).*

*Certain basic inadequacies of the models are described. Lee's model lacks a sufficiently broad assignment of probabilities to events of interest, with the result that the blocking probability is improperly defined; at the same time it bases congestion formulas on network conditions never achieved in practice. The thermodynamic model deals only with genuine system states, but makes calling rates depend unrealistically on available paths. Neither the graph model as originally proposed, nor the thermodynamic model, can take into account routing procedures. The author's Markov model is free of these drawbacks, but at this price: in nearly all practical situations in which losses occur, it leads to hitherto insurmounted combinatorial and computational difficulties.*

*To stress and illustrate the effect that routing has on loss, the blocking probability formulas of all three models are compared at low traffic: it often turns out that when the first two models indicate that (with $\lambda$ = offered traffic) loss $= O(\lambda^m)$, $\lambda \to 0$, an analysis based on routing shows that in fact loss $= o(\lambda^m)$, $\lambda \to 0$.*

## I. INTRODUCTION AND SUMMARY

In recent years several stochastic models for random traffic in large telephone connecting networks have been proposed. In 1955, C. Y. Lee[1] described what has come to be known as the "probability linear graph" model, an outgrowth of earlier work of Kittredge and Molina. In 1957, P. Le Gall[2] developed and used essentially the same model as Lee. In

1963 the author presented another model,[3] which he called "thermodynamic" because of its resemblance to statistical mechanics, and a third one[4] that was formulated in an effort to take routing methods into account and to meet certain drawbacks of the thermodynamic model.

These three models form a progression of decreasing simplicity and increasing realism, and they exhibit the trade-off between verisimilitude and computational difficulty: the more realistic the model, the harder it is to calculate anything in it. Their existence also affords a limited opportunity for trying to make comparisons, e.g., to see in what ways the various assumptions made affect the formula for blocking probability.

Our object here is to discuss the respective inadequacies of all three models, to compare their blocking probability formulas at low traffic, and to stress and illustrate the point that routing has a basic structural effect on the probability of loss: in particular, it often turns out that when the first two models indicate that, with $\lambda$ = offered traffic, loss $= O(\lambda^m)$ as $\lambda \to 0$, an analysis actually based on reasonable routing shows that loss $= o(\lambda^m)$, $\lambda \to 0$.

## II. DISCUSSION OF LEE'S "PROBABILITY LINEAR GRAPH" METHOD

The 'probability linear graph" approach to the study of connecting networks has been extensively described by its proposer C. Y. Lee,[1] and more recently by R. F. Grantges and N. R. Sinowitz.[5] Therefore, we include only the following résumé of the method: to calculate the congestion incurred by traffic between an inlet $u$ and an outlet $v$, attention is focussed on the graph $G$ defined by the possible (i.e., permitted) paths through the network from $u$ to $v$; $G$ consists of all nodes and branches through which some path from $u$ to $v$ passes. Naturally, $G$ is connected. Given any assignment of occupancies to the branches of $G$, i.e., any (complete) specification of which branches of $G$ are busy and which are idle (at a particular juncture of network operation), it is possible to examine $G$ to see if there is a path from $u$ to $v$ (no branch of which is busy). The method now assigns a probability distribution to the possible occupancies by postulating that a link $l$ of $G$ is busy with probability $p_l$ independently of all other links. The congestion for $u$ and $v$ is then calculated as the probability that this distribution assigns to the event "There is no path from $u$ to $v$". The probabilities $\{p_l, l$ a link of $G\}$ are chosen to reflect the loads carried by the links in the network.

A foremost merit of the Molina-Kittredge-Lee proposal is of course that it provides a simple way of at least approaching the massive problem of theoretically determining the grade of service in a connecting network. For small networks the calculations can be done by hand, and for large ones, in which the graph $G$ is complex, it is feasible to use computers to evaluate the polynomials that arise,[5] or to use approximations.[6] For these reasons alone Lee's model merits serious consideration. Indeed, R. F. Grantges and N. R. Sinowitz claim: "The utility of the results obtainable from Lee's model is well known. When the specific [average] branch occupancies are chosen rationally, the calculated blocking agrees well enough with real blocking figures (obtained from full-scale simulation or measurement) for many engineering and design purposes." (Ref. 5, p. 977.) However, these same authors, while using Lee's model, also state explicitly: "Unfortunately, in complex switching networks, substantial differences may exist between the estimates obtained with Lee's analytical technique and actual performance determined by field measurement or full-scale (complete) simulation." (Ref. 5, p. 1000.) It appears, then, that some discussion and evaluation of the basis of Lee's model might help to indicate where and why it departs from reality.

It is often stated (e.g., Ref. 5, p. 969) that a principal unrealistic feature of Lee's model is the assumption that the occupancies of the links are statistically independent. However, the Kittredge-Molina-Lee approach involves some problems most of which are independent of this assumption:

(*i*) It does not assign probability to enough events of interest, so that there is difficulty in properly defining the probability of blocking.

(*ii*) It may assign substantial probability to (and base congestion calculations on) events which can never occur in real life. (This fact depends, of course, on the independence assumption.)

(*iii*) It does not recognize (wide or strict sense) nonblocking networks.

(*iv*) It does not take into account the effects of routing decisions.

Noting most of these difficulties, Grantges and Sinowitz[5] have devised ingenious modifications of Lee's basic method in order to meet them, and to increase the realism of the graph model. When each procedure is compared against full-scale simulation, these refinements give a remarkable improvement in accuracy over Lee's original proposal. The modifications are suitable for computer simulation of Lee's method, and do not give rise to a formula for analysis; thus, we are not able to include them in the low traffic comparison at the end of this paper.

Problems (*i*) and (*ii*) are discussed in the next two sections; (*iii*) and (*iv*) are considered after a description of the thermodynamic model, which also has such drawbacks.

## III. INCOMPLETENESS

By a full-fledged stochastic model for network operation, we mean one in which every event observable in the real-life system has a counterpart in the model which is assigned a probability. If this seems an excessive requirement, let us agree that at least any event depending on the busy or idle condition of crosspoints and links in the network is to be assigned probability. It is a pertinent comment that Lee's calculation is not based on such a model, not even in the weak sense agreed on.

The incomplete character of the assignment of probability in Lee's model has serious consequences. For example, not all events depending on what inlets or outlets are busy or idle are assigned probability. In particular, when the congestion for traffic from inlet $u$ to outlet $v$ is under consideration, the model does not assign a probability to the event.

$$\{\text{the call from } u \text{ to } v \text{ is blocked}\},$$

in the customary sense of 'blocked'. This is because it may be true, at a particular moment, that there is no path from $u$ to $v$ on $G$ in the sense that every possible path from $u$ to $v$ has at least one busy link, while $u$ or $v$ or both may be busy talking to other terminals over other links. In such a case, we would not say that a $u$, $v$ call was blocked; only if there was no path and *both* $u$, $v$ were *idle* would we say they were blocked. However, the event that $u$ is idle, or that $v$ is idle, is not (and cannot be) considered, since no event of this form has been assigned probability. Lee's model assigns probability to so few events that it cannot distinguish between the above two cases. Indeed, this circumstance is directly responsible for the model's inability to recognize a nonblocking network when it sees it. In such a network the event

$$\{\text{every path from } u \text{ to } v \text{ has at least one busy link}\}$$

will in a reasonable stochastic model have positive probability, but the event

$$\{\text{the call from } u \text{ to } v \text{ is blocked}\}$$

has probability zero. It is in part because Lee's model calculates the probability of the former event that it gives the wrong answer for nonblocking networks; to the latter event it does not even assign probability.

The problem just discussed has been treated by Grantges and Sinowitz[5] in their prefatory remarks, and at considerable length by E. Wolman.[7]

## IV. IRRELEVANT STATES

The probability linear graph model not only fails to assign probability to events like blocking which should have it; it also does assign it to events which never occur in an operating exchange. It bases congestion calculations on situations, i.e., conditions of the network, which never arise in practice. Moreover, these irrelevant situations can be so numerous as to greatly outnumber the relevant ones. The model assigns these irrelevant situations probabilities that are comparable to those assigned to the relevant states. The applicability of any calculation depending so heavily on irrelevant material is open to question: it is very hard to see why these irrelevant states do not swamp the ones of real interest.

To illustrate our point about irrelevant situations, suppose that for some inlet $u$ and outlet $v$ the graph $G$ determined by the paths from $u$ to $v$ includes every inlet and outlet on some square switch, deep in the middle of the network. Now, in every physically meaningful state of the network, reachable under normal operation, this switch will have as many idle inlets as outlets. However, in this case, Lee's method will also base congestion on situations with $m$ inlets busy, $n$ outlets busy, and $m \neq n$. It is easily seen that these are in the vast majority, and that Lee's model assigns them probabilities comparable with those assigned to situations with $m = n$. (The second of these facts depends, of course, on the "independence of links" assumption.)

Assessing the effect of the irrelevant states is difficult, but their presence may help to explain the variable agreement of Lee's model with experiment: when the proportion of blocking states is the same in the set of relevant states as in that of all states, the model may be accurate; when the inclusion of irrelevant situations produces bias—either too low or too high a proportion of blocking states in the set of all states—the model is inaccurate.

## V. DISCUSSION OF THE THERMODYNAMIC MODEL

The thermodynamic model[3] for equilibrium traffic in a telephone connecting network is obtained as follows: the physically meaningful states of the network are collected in a partially ordered set $(S, \leqq)$, and a distribution $\{q_x, x \,\varepsilon\, S\}$ of probability over $S$ is defined by the condition that $q$ maximize the entropy functional

$$H(q) = -\sum_{x \varepsilon S} q_x \log q_x$$

subject to the condition that $\sum_{x \varepsilon S} |x| q_x =$ carried load (a given number).

With $\Phi(z) = \sum_{x \epsilon S} z^{|x|}$ and $\xi$ the positive solution of

$$\text{carried load} = \xi \frac{d}{d\xi} \log \Phi(\xi), \tag{1}$$

$q$ has the form

$$q_x = \Phi^{-1}(\xi)\xi^{|x|}. \tag{2}$$

An extensive discussion of the thermodynamic model is given in Ref. 3. We confine ourselves here to a brief presentation of points relevant to comparing it with Lee's model, and with that of Ref. 4.

(*i*) It provides a full-fledged stochastic model for traffic in the network: each possible meaningful state is assigned probability in a simple way [formula (2)]. This great advantage is of course obtained at the considerable price of introducing the complicated set $S$ of states; for many purposes, calculation with $S$ can be replaced by calculation with the numbers $| L_n |$, where $L_n$ is the set of states with $n$ calls in progress, and $| \cdot |$ indicates cardinality. While this replacement is an enormous simplification over use of $S$, the determination of the numbers $| L_n |$ is nevertheless a formidable and unsolved problem; however, it is also one on which virtually no effort has been expended except in unpublished work of A. J. Goldstein.

(*ii*) In order to construct a realistic model, it is not enough to take into account all and only the meaningful states in some full-fledged stochastic model. It is also necessary that the model be based on a realistic description of the rates at which the system moves from state to state. In this second respect the thermodynamic model falls quite short. It was pointed out in Ref. 4 that the thermodynamic model corresponds closely to random choices of routes for calls, together with the artificial feature that the calling-rate of a call depended on the number of paths available for it.

(*iii*) The thermodynamic model shares with Lee's the drawback that it is incapable of describing the effect of general routing policies. It is known that these effects can change substantially the probability of blocking, in some cases by a factor of ten.[8] The reason for this is, roughly, that proper routing largely avoids the disastrous states in which many calls are blocked. Oblivious of routing, the thermodynamic model gives positive probability to any path on $(S, \leqq)$, the "bad" paths receiving probabilities comparable to those of the "good" ones.

(*iv*) Since the thermodynamic model assigns probability to every meaningful state, it is possible to give a reasonably satisfactory definition of blocking probability in it. In analogy with Ref. 4 we define it as

$$\Pr\{bl\}_\theta = \frac{\sum_{x \varepsilon S} \xi^{|x|}\beta_x}{\sum_{x \varepsilon S} \xi^{|x|}\alpha_x}, \tag{3}$$

where $\xi$ is as in (1), $\theta$ stands for 'thermodynamic',

$\beta_x$ = number of idle inlet-outlet pairs in $x$ that are blocked,
$\alpha_x$ = number of idle inlet-outlet pairs in $x$.

This formula holds an advantage over Lee's in that it gives the value zero for a strictly nonblocking network. However if the network is only nonblocking in the wide sense, i.e., if it is only nonblocking if the right routing is used, then the blocking as given by (3) will not distinguish this behavior: it will give a positive answer that does not depend on how routing is actually carried out.

## VI. LOW TRAFFIC BEHAVIOR: CALIBRATION

The differences between the three models being considered here are particularly evident when they are used for studying a network that is nonblocking, whether strictly or in the wide sense. Needless to say Lee's method cannot distinguish the nonblocking behavior at all, while the thermodynamic model can recognize a strictly nonblocking network ($\beta_x \equiv 0$), but cannot distinguish lack of blocking due to proper routing (nonblocking in the wide sense.)

In an effort to provide an analytical comparison between Lee's model, the thermodynamic model, and the Markov process model of Ref. 4, we shall examine the leading terms of the blocking probability formula in each model for low traffic in a Clos 3-stage network with $r$ $n \times m$ outer switches, $m$ $r \times r$ middle switches, $N = rn$ inlets (outlets), and $n \leqq m \leqq 2n - 2$. This network is depicted in Fig. 1.

Such a comparison can only be sensible if the link occupancies in the models agree asymptotically. In many ways it would be more desirable to calibrate by requiring the same carried load in each model; but in Lee's model the only way of defining this requirement is by reference to the link occupancies, a procedure equivalent to ours.

It is to be kept in mind that the comparison to be made is carried out on the basis of asymptotic formulas valid only for sufficiently small values of $\lambda$. The range of loss probabilities over which the comparisons are performed is not known and could conceivably fall entirely outside the domain of practical relevance.

In the model of Ref. 4 the inlet occupancy is reasonably defined as

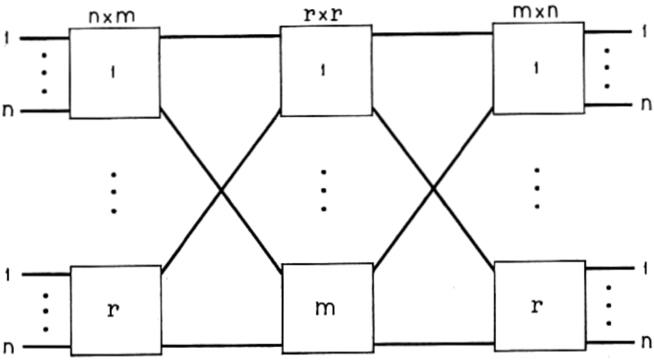

Fig. 1 — 3-stage Clos network.

$$q = \frac{\text{carried load}}{N}.$$

Since it is elementary that

$$\text{carried load} = \lambda N^2 + o(\lambda), \qquad \lambda \to 0$$

it is reasonable that the link occupancy $p$ in Lee's model be

$$p = \frac{\text{carried load}}{N} \times \frac{n}{m} = \frac{\lambda r n^2}{m} + o(\lambda), \qquad \lambda \to 0.$$

With this calibration the blocking in Lee's model for the network of Fig. 1 is

$$\begin{aligned}[1 - (1 - p)^2]^m &= (2p)^m + o(p) \\ &= \lambda^m \left(\frac{2rn^2}{m}\right)^m + o(\lambda), \qquad \lambda \to 0.\end{aligned}$$

For the thermodynamic model the calibration is a little more involved. The basic requirement is that the parameter $\xi$ used in that model satisfy

$$\text{carried load} = Nq = \frac{\sum_{x \varepsilon S} \xi^{|x|} \mid x \mid}{\sum_{x \varepsilon S} \xi^{|x|}}.$$

It is easy to see that the ratio has the form

$$\mid L_1 \mid \xi + \Psi(\xi), \mid L_1 \mid = \text{number of states with one call in progress,}$$

with $\Psi(\xi) = o(\xi)$ as $\xi \to 0$. Since for complex $z$ small enough

$$\frac{|\Psi(z)|}{|L_1|} \leqq \left| z - \frac{Nq}{|L_1|} \right|$$

it follows from Lagrange's theorem[9] that for $q$ small enough,

$$\xi = \frac{Nq}{|L_1|} + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{d^{n-1}}{dx^{n-1}} \left( \frac{\Psi(x)}{|L_1|} \right)^n \Bigg|_{x=Nq/|L_1|},$$

$$\xi = \frac{Nq}{|L_1|} + o(q) = \frac{\lambda N^2}{|L_1|} + o(\lambda), \qquad \lambda \to 0.$$

It is easily verified that $|L_1| = nrmrn = N^2 m$, whence

$$\xi = \frac{\lambda}{m} + o(\lambda), \qquad \lambda \to 0. \tag{4}$$

(This formula exhibits the sense in which at equal carried loads the parameter $\xi$ of the thermodynamic model is about $1/m$ times the parameter $\lambda$ as a result of the increased calling-rate in that model due to its proportionality to the number of available routes.)

## VII. LOW TRAFFIC BEHAVIOR: THE EFFECTS OF ROUTING

It can be seen that, for the Clos network under discussion here, $\alpha_0 = N^2$, so that (3) and (4) give

$$\Pr \{bl\}_\theta = N^{-2} \frac{\lambda^m}{m^m} \sum_{|x|=m} \beta_x + o(\lambda^m).$$

In the model of Ref. 4, the blocking probability for this same network has the form

$$\Pr \{bl\} = N^{-2} \frac{\lambda^m}{m!} \sum_{|x|=m} r_x \beta_x + o(\lambda^m),$$

where

$\beta_x$ = number of calls blocked in state $x$
$r_x$ = number of ways of ascending from 0 to $x$ along paths permitted by the routing rule in use.

We thus arrive at three low traffic formulas for loss all expressed in terms of $\lambda$: as $\lambda \to 0$,

$$\Pr\{bl\}_\theta = N^{-2}\frac{\lambda^m}{m^m}\sum_{|x|=m}\beta_x + o(\lambda^m), \tag{5}$$

$$\Pr\{bl\}_{\text{Lee}} = \left(\frac{2rn^2}{m}\right)^m \lambda^m + o(\lambda^m), \tag{6}$$

$$\Pr\{bl\} = N^{-2}\frac{\lambda^m}{m!}\sum_{|x|=m}\beta_x r_x + o(\lambda^m). \tag{7}$$

The sums in these formulas are in general not easy to calculate, depending as they do on network structure and routing. Our point, though, is precisely that the dependence on *routing* is crucial, since by making the bad states relatively inaccessible we can make the sum

$$\sum_{|x|=m} r_x\beta_x$$

small, even to the point of being zero. In such a case the first two blocking formulas do not even have the right leading term.

To see how this comes about, we refer to Fig. 2, which shows a typical blocking state of dimension $m$ of the Clos network of Fig. 1. It is clear that if some of the calls were to double up on the middle switches, in-

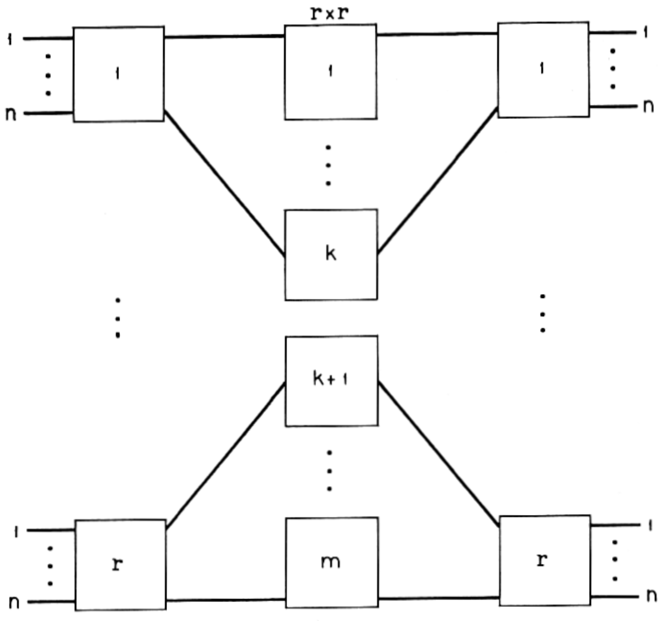

Fig. 2—Typical state with dimension $m$ and some blocked calls.

TABLE I

| Network parameter | | | Coefficient of $\lambda^m$ in blocking formula | | |
|---|---|---|---|---|---|
| $n$ | $m$ | $r$ | Lee | $\theta$ | Ref. 4 |
| 2 | 2 | 2 | 64 | 4 | 0 |
| 3 | 3 | 2 | 1728 | 27 | 0 |

stead of wastefully each using a middle switch, there need be no blocking. Now for $r = 2$, $m \geqq [3n/2]$ the network under study is nonblocking in the wide sense. That is, these conditions on $r$, $m$, and $n$ ensure that as long as correct routing decisions are made, no call need be blocked. The "correct" routing that achieves this performance consists precisely in not using an empty middle switch when a partly-filled one is available. It is highly likely, as the results of Ref. 8 suggest, that this advice is good even when $r > 2$, $m < [3n/2]$. Let then $R$ be a routing matrix for the network of Fig. 1 which embodies the above advice. It can be seen that

$$r_x = (R^{|x|})_{0x} = 0$$

for $| x | = m$ and $\beta_x > 0$. I.e., all the blocking states of dimension $m$ are unreachable from 0 in $m$ steps under $R$, because $R$ insists that empty middle switches be used only when partly-filled ones are unavailable. This very reasonable routing makes the coefficient of $\lambda^m$ in Pr {bl} vanish.

Table I shows the coefficient of $\lambda^m$ in the low traffic formulas (5) to (7) for two very small networks; for (7) it has been assumed that no unblocked call is rejected and that routing is optimal, i.e., minimizes loss.

## VIII. CONCLUSION

This paper has provided one more illustration of a situation that traffic experts are well aware of, namely, that blocking probabilities in connecting networks can be computed only under assumptions that are not satisfied in practice. There is concrete evidence indicating that results obtained within the framework of such approximate models can be of practical value. By considering the respective advantages and drawbacks of a spectrum of models, it is possible to discern to some extent the effect of the various assumptions made on the structure of the formula for loss.

## IX. ACKNOWLEDGMENT

The author is indebted to A. Descloux and E. Wolman for constructive suggestions.

## REFERENCES

1. Lee, C. Y., Analysis of Switching Networks, B.S.T.J., *34*, November, 1955, pp. 1287–1315.
2. Le Gall, P., Méthode de Calcul de l'Encombrement dans les Systèmes Téléphoniques Automatiques à Marquage, Ann. Télécomm., *12*, 1957, pp. 374–386.
3. Beneš, V. E., A "Thermodynamic" Theory of Traffic in Connecting Networks, B.S.T.J., *42*, May 1963, pp. 567–607.
4. Beneš, V. E., Markov Processes Representing Traffic in Connecting Networks, B.S.T.J., *42*, November, 1963, pp. 2795–2837.
5. Grantges, R. F. and Sinowitz, N. R., NEASIM: A General-Purpose Computer Simulation Program for Load-Loss Analysis of Multistage Central Office Switching Networks, B.S.T.J., *43*, May, 1964, pp. 965–1004.
6. Lee, L. and Brzozowski, J. A., An Approximate Method for Computing Blocking Probability in Switching Networks, IEEE Trans. Commun. Tech., *COM-14*, April, 1966, pp. 85–93.
7. Wolman, E., On Definitions of Congestion in Communication Networks, B.S.T.J., *44*, December, 1965, pp. 2271–2294.
8. Beneš, V. E., Programming and Control Problems Arising from Optimal Routing in Telephone Networks. Abstract in SIAM J. Control, *4*, February, 1966, pp. 6–18. Text in B.S.T.J., *45*, November, 1966, pp. 1373–1438.
9. Whittaker, E. T. and Watson, G. N., *Modern Analysis*, 4th Ed. Cambridge, at the University Press, 1948, p. 133.