*A teleprocessing system may include many low-speed terminals at great distances from the computing center. Specification of a communication network for connecting the remote terminals to the central computer constitutes an important design problem.*

*An iterative method for obtaining an approximate solution to an optimum network is presented. The method assumes that an acceptable line utilization factor is given.*

# On teleprocessing system design

## Part II A method for approximating the optimal network

**by L. R. Esau and K. C. Williams**

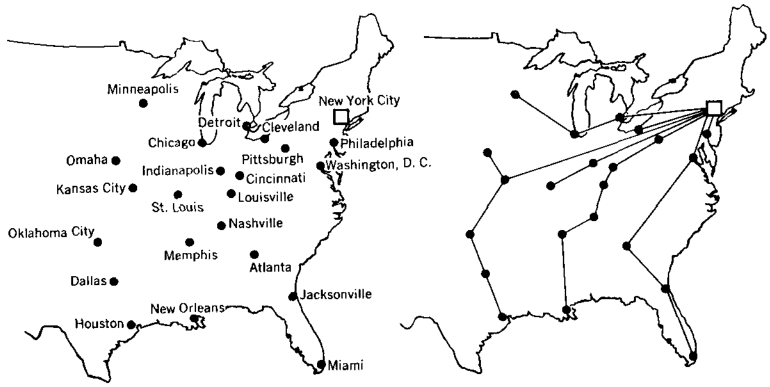
For practical purposes, the analyses involved in designing a teleprocessing system can be divided between studies pertaining to the design of the systems center and those pertaining to the communication network. The latter studies, in turn, can be subdivided into analyses of both acceptable line loadings and optimal line configurations.

The problem of determining acceptable line loadings is one that depends heavily upon the character of the application. If the response-time requirement is somewhat lenient, line loadings may be satisfactorily estimated on the basis of rules of thumb from comparable systems or from observations on pilot systems. If the response-time requirement is less lenient, on the other hand, the estimation of acceptable line loadings may merit the development of a complex probability model or the use of extensive simulation exercises.[1]

Regardless of the method used to obtain acceptable line loadings, it is necessary to determine a network configuration that connects terminals with the control center. Because lines have a significant affect on the cost of most teleprocessing systems, it is important to find a configuration that is reasonably optimal. Experience suggests that the operating cost of a network can often be substantially reduced by an initial investment in a configuration analysis.

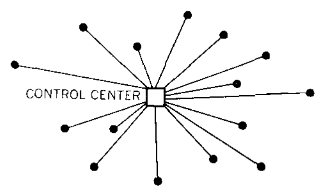
Although discrete linear programming approaches for obtaining the optimum design have been described, simplicity and

Figure 1 Network configuration




computational ease are often more important than rigor in the iterative give-and-take milieu of system design.[2] We therefore discuss a straightforward approximation that requires a modest amount of processing. While the method is too tedious for manual solutions, it lends itself well to computer computation. With a medium-sized computer, problems involving, say, two hundred terminals can be solved in under ten minutes. Experience on test problems suggests that the approximations obtained are sufficiently accurate for engineering practice; in fact, for problems of nontrivial size, manual checks have always failed to improve on solutions. The method assumes that suitable line loading factors are given.

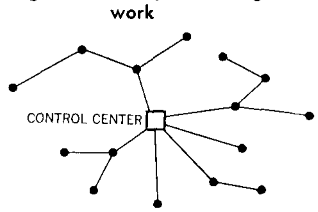
Figure 2 Point-to-point network




## An algorithm for constructing a multipoint network

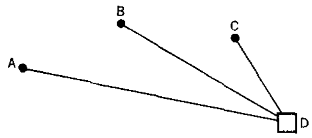
Figure 1 illustrates a typical network consisting of remote locations, communication facilities, and a single control center (data processing center). For given performance requirements, the objective of network design is to determine an economical pattern for connecting remote locations to the center. The size and type of network required depends upon the application. The simplest type of network is the one in which a control center is connected to each remote location by a separate line (see Figure 2). Inasmuch as there is only one way in which to make the connections, a network of this type presents no configuration problem. But by its very nature, it tends to be the most expensive type. Such a network can seldom be justified except in specialized applications where terminals are very heavily used or equipment design precludes the sharing of lines.

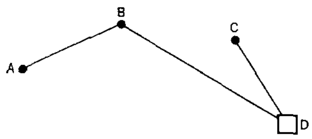
Figure 3 Multipoint linkage network

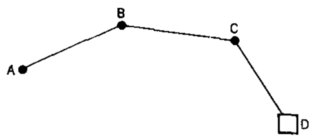



A second network type retains the single control center, but is characterized by the presence of one or more *multipoint* lines. In each multipoint line, the link that feeds the control center will be termed the *central link*. For example, the configuration of Figure 3 contains four multipoint lines and two single link lines.

Figure 4



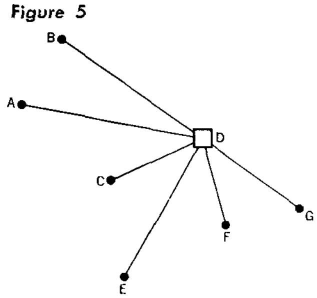

A network that contains multipoint lines will be referred to as a *multipoint network*. A multipoint network can be difficult to optimize—even with a relatively small number of remote locations—because the number of possible configurations that can exist is very large. The magnitude of a multipoint configuration problem for $n$ points can be illustrated by the fact that the total number of possible configurations is considerably greater than $n$ factorial. For example, when $n = 5$, the number of possible configurations reaches 1416. Moreover, when $n = 10$, the number of combinations exceeds three and one half million. Thus, even for modest values of $n$, it is unrealistic to try to evaluate all possible configurations.

In order to construct a communications network, some means of establishing the distance between points is necessary. This is generally done with the aid of a rectangular grid system, each point being identified by a unique pair of coordinates. The distance between two points $P_1$ and $P_2$ can then be defined as a function of the coordinates $(x_1, y_1)$ of points $P_1$ and $(x_2, y_2)$ of point $P_2$. The most widely-used grid system for the continental United States, Canada, and Mexico provides $V$-$H$ (vertical and horizontal) coordinates for use in determining distance.[3] The distance between two points $P_1$ and $P_2$ with coordinates $(V_1, H_1)$ and $(V_2, H_2)$, respectively, is defined as

$$d = \sqrt{\frac{(V_1 - V_2)^2 + (H_1 - H_2)^2}{10}}$$

The factor 10 enters because of the manner in which the coordinates are scaled in this grid system.

Figure 5




If a single-line tariff applies, and if cost is proportional to distance, distance will suffice as the basic variable in the construction of a most-economical configuration. On the other hand, if more than one tariff applies, e.g., an interstate tariff plus one or more intrastate tariffs, cost must be taken into account in constructing an optimum configuration.

Since a communication line is capable of transmitting only a limited amount of traffic in a given interval of time and still satisfy the application requirements, the desired configuration will be determined, in part, by the amount of traffic at each remote location and the amount the line can handle. The method to be described assumes that the line-loading capabilities are given.

In explaining the algorithm, we will assume that cost is proportional to distance; the more general case is discussed later. The algorithm assumes that a maximum-distance configuration is one in which each remote location is connected to the center by a single link. (If the traffic at a location exceeds the capacity of a single line, the required number of lines is determined.) Each initial fully-loaded line is set aside because it obviously cannot be multipointed. Traffic volumes permitting, however, the formation of multipoint lines for the remaining points may lead to distance savings. For example, as illustrated in Figure 4, a savings equal to the difference between the lengths of the central link A and link AB can be realized if points A and B are multipointed on line B. A still greater savings could be achieved by placing all three points A, B, and C on one line (by forming link BC and removing central link B). In general, if a savings is possible, it is attained by removing a central link and inserting another link. Such changes are made one at a time, and any change can affect subsequent changes.

To determine the order in which changes should be made, let $T_{X:YZ}$ denote the *trade-off value* involved in removing central link $X$ and forming link $YZ$. For each line in a configuration, there may be several $T$'s, one for each line to which the given line can be connected and still perform satisfactorily. Since combined traffic cannot be greater than one line can handle, traffic volume is checked before an attempt is made to compute $T$.

Assume that Figure 5 represents an initial maximum cost configuration, and that fully loaded lines have been set aside. Let each line be identified by the point that is connected directly to the center. Initially, there are as many lines as remote points, but, as multipoint lines are formed, the number will decrease. Let all of the $T$'s be computed for a given line, and let $\hat{T}$ denote their maximum. Moreover, let $\hat{\hat{T}}$ denote the maximum for all lines. If $\hat{\hat{T}} = T_{A:AB}$, then the configuration in Figure 5 would be modified as shown in Figure 6.

Removal of a central link may render certain $T$'s invalid. For example, if $\hat{T}$ for line B is given as $T_{A:AB}$, this is no longer valid if central link A does not exist. The maximum $T$ for line B, when initially determined, may have indicated replacing central link B by link BC. Since line B now has the combined traffic of points A and B (see Figure 6), line C may not be able to handle the total traffic. If this is the case, the $\hat{T}$ for line B is no longer valid and a new one must be computed. After the $\hat{T}$'s are updated as necessary, another $\hat{\hat{T}}$ is chosen. If $\hat{\hat{T}}$ turns out to be $T_{E:EC}$, then the configuration in Figure 6 is modified as shown in Figure 7.

It should be noted that this procedure provides for the building of one or more multipoint lines at a time. A point on a line is always linked to its nearest neighbor as a result of changes. In Figure 7, the set of $\hat{T}$'s consists of four values, one each for lines B, C, F, and G. If any of the initial $\hat{T}$'s for these lines has been invalidated due to changes, they must be recomputed. However, a new $\hat{T}$ is never greater than the old one. Therefore, until an invalid $\hat{T}$ is chosen as a $\hat{\hat{T}}$, it need not be recomputed.

When computing $T$ for line B relative to line C, all points in line C must be considered. However, by definition of trade-off value, all points have already been connected to their nearest acceptable neighbor. Thus, the linkages previously established for lines B and C will remain; it is only necessary to consider the shortest link that will connect the two lines. The difference between this link and central link B will produce the $T$ for line B relative to line C. Of course, if line C cannot handle the additional

Figure 6

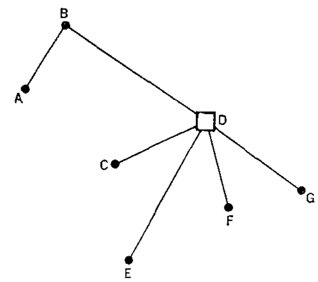



Figure 7

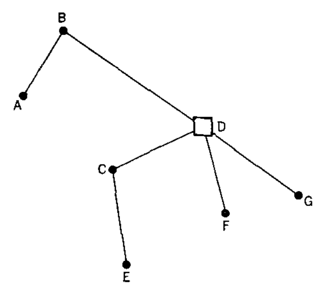



Figure 8

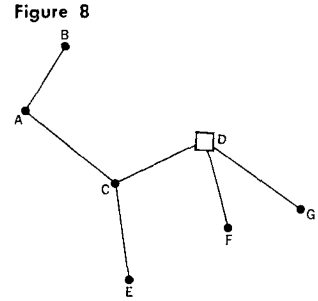



Figure 9

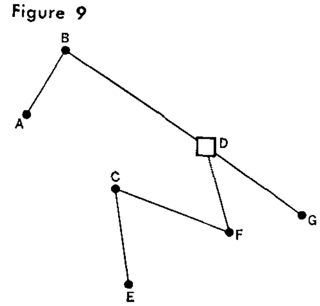
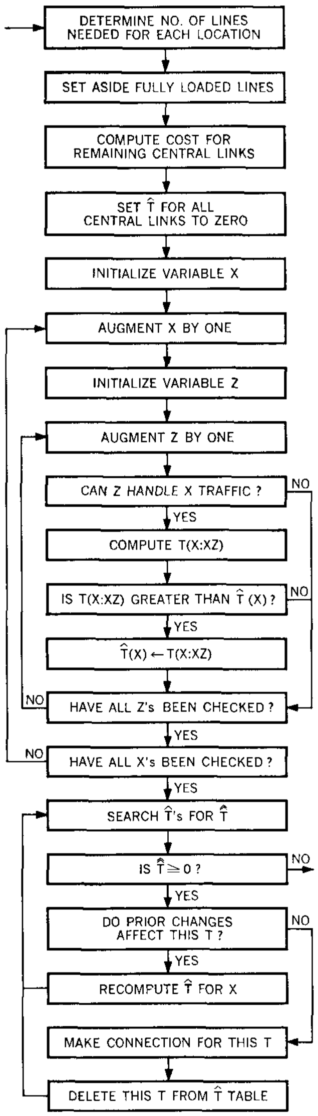


Figure 10 Flow diagram of basic method




traffic from line B, it should not be considered in computing a new $T$ for line B. Assuming that C can handle the traffic, and that it produced the greatest $T$, such that $\hat{\hat{T}} = T_{B:AC}$, Figure 7 would be modified as shown in Figure 8. If, on the other hand, $\hat{T}$ for line C became $\hat{\hat{T}}$ and implied the replacing of central link C with link CF, Figure 7 would be modified as illustrated in Figure 9. The required configuration has been formed when $\hat{\hat{T}}$ is found to be negative.[4] A flowchart of the method is shown in Figure 10.

## Extensions to the basic algorithm

If distance and cost are not proportional, the algorithm can still be made applicable by basing the $T$'s on cost rather than distance. Although cost is often proportional to distance, there are cases in which the shortest distance does not imply the least cost. This can be illustrated by a network wherein both interstate and intrastate tariffs apply. Since the cost per mile is generally different for the two tariffs, it becomes necessary to determine not only distance but also which tariff to use for costing purposes. A line is classified as intrastate if and only if all links forming that line are within the same state. When one of the links extends across the state boundary, the line becomes interstate and all links are costed according to the interstate tariff. Since intrastate tariffs are generally higher in cost, it is possible that the maximum $T$ for a given line does not result from connecting a point to its nearest neighbor.

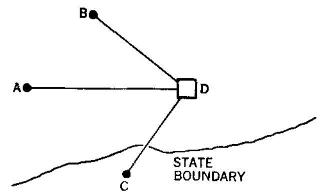
Assume in Figure 11 that points A and B lie within the same state as the center point D, but point C lies outside the state. In computing $\hat{T}$ for line A, two $T$'s, one associated with line C and another with line B, are evaluated. It is obvious, in the case of distance, that replacement of central link A with link AB would produce a greater savings than replacing A with AC. However, since the link AC will be costed on the basis of the interstate tariff, whereas AB is based on the intrastate tariff, the replacement of A with AC may produce the greater savings.

An important difference between implementation of the minimum-distance and the minimum-cost algorithm is that whenever a $T$ is computed it cannot be assumed that each point has been connected to its nearest neighbor. It is therefore necessary to restructure the linkages to assure that each point is linked to its nearest neighbor before evaluating costs to determine $T$.

In the basic algorithm, it is assumed that only one line type is involved. The amount of traffic that a line can handle is a function of transmission speed and consequently line type; the greater the speed of a line, the larger the volume of traffic it can handle. However, since the cost per mile increases with an increased speed capability, it is not obvious which type of line will form the most economical configuration. By using the algorithm to construct configurations for each type line to be considered, the most economical type can then be selected.

Although the cost per mile usually increases as transmission capability increases, it increases at a slower rate. Therefore, a higher speed line will provide more capability for the same amount of money. But, since the geographical relationship of the remote points and their respective traffic volumes influence the decision to link various points of the same line, lines may be formed that are not loaded to their full capacity. The only way to determine if higher speed transmission will be more economical for a given application is to construct configurations for alternate speeds and compare their costs.

Figure 11




In some applications, a combination of lines of different speeds will produce the most economical network configuration. By forming a configuration for the highest speed that is applicable, a close examination of the lines formed will indicate which lines are loaded near their maximum capability. These lines are, in general, the most economical for connecting the points involved to the center. Lines that are lightly loaded may be more economically structured using lower speed lines. All points associated with such lines can be reconfigured based on a lower speed capability and costing structure. The resulting cost can be compared with the higher speed line cost and the most economical used.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of Mr. Wei Chang in the development of the algorithm discussed in this article.

CITED REFERENCE AND FOOTNOTES

1. See Seaman, Part VI, on the general subject of simulation. We know of no reference to a standardized queuing model specifically applicable to the line-loading problem.
2. The reader who is interested in more general approaches to network design may consult W. H. Kim and R. T. Chien, *Topological Analysis and Synthesis of Communication Networks*, Columbia University Press, New York (1962). A different computational approach is discussed in Harry M. Markowitz and Alan S. Manne, "On the solution of discrete programming problems," *Econometrica* **25**, No. 1, 84–110 (January 1957).
3. *A List of Rate Centers and Control Offices*, Tariff FCC #255, filed by American Telephone and Telegraph Company.
4. Zero trade-off values do not lead to savings but have the virtue of forcing the configuration to a minimum number of lines.