CONCEPTS OF OPTIMALITY IN ALTERNATE ROUTING NETWORKS

R. J. Harris
University of Adelaide
Adelaide, South Australia

1. INTRODUCTION

One important aspect of planning a telephone network is the determination of the optimal number of circuits to be provided between pairs of exchanges. One of the difficulties which must be overcome in this planning process concerns the question of what constitutes an "optimal" design. The following is a list of possible interpretations of the word "optimal".

(i) The network should be designed in such a way that the total cost is a minimum and that traffic losses are taken to be a predetermined fraction $B^k$ of the traffic offered between origin-destination pairs $k$.

(ii) Planners should attempt to minimize the total cost of traffic passing from an origin to a destination. (Traffic losses are again taken to be a predetermined fraction of the traffic offered between the origin-destination pair.)

(iii) The network should be dimensioned in such a way that routes which are carrying traffic have a maximum traffic efficiency per unit circuit cost. (Origin-destination losses being fixed.)

(iv) The grade of service on last choice links is a predetermined value, and the total network cost is to be minimized.

The concepts (i) and (iv) have been proposed by Rapp [10] as possible criteria for obtaining "optimal" telephone network designs. The first concept has been used as a basis of a dimensioning procedure by Berry [1,2], and (iv) has been used by Pratt [7], and Wallström [8] as a basis for their dimensioning models.

After defining the symbols to be used (Section 2) in this paper, a mathematical description of the optimality requirements for the first three optimizing principles will be given (Section 3). In Section 4 a simple example will be given which illustrates the differences between the first three concepts listed above. Sections 5 and 6 will discuss two methods for dimensioning alternate routing networks in order to demonstrate how the various optimizing principles can be applied.

2. SYMBOLS AND DEFINITIONS

Consider a mathematical network $T$ which consists of a set $N$ of nodes and a set $L$ of $t$ directed links. In this network we specify a set of $n$ Origin-Destination pairs and denote them by $O^k \rightarrow D^k$, $k = 1, \ldots, n$ and a set of $m_k$ allowable chains joining each pair, which will be denoted by $m_k^i$, $j = 1, \ldots, m_k$. Let $f^k$ be the offered traffic destined for $D_k$ from $0^k$. The flow on chain $m_k^i$ will be denoted by $h_{ij}^k$ and the total flow $f_i$ on link $i$ is

$$ f_i = \sum_k a_{ij}^k h_{ij}^k \quad 1 = i, \ldots, t \quad (1) $$

where $a_{ij}^k = \begin{cases} 1 & \text{if link } i \text{ is on chain } m_k^i, \\ 0 & \text{otherwise.} \end{cases}$
A chain flow pattern \( h \) will be defined as
\[
    h = \begin{bmatrix}
        h^1 \\
        \vdots \\
        h^k \\
        \vdots \\
        h^n
    \end{bmatrix}
\]
where
\[
    m^k = \begin{bmatrix}
        h_1^k \\
        \vdots \\
        h_{m_k}^k
    \end{bmatrix}
\]
A link flow pattern \( f \) will be defined by
\[
    f = \begin{bmatrix}
        f_1 \\
        \vdots \\
        f_s \\
    \end{bmatrix}
\]
In addition to the above notation the following symbols will be used:
\( i \) symbol used to indicate an arbitrary link.
\( k \) symbol used to denote an arbitrary OD pair.
\( j \) symbol used to denote an arbitrary OD pair.
\( M_k \) the mean of the traffic offered to link \( i \).
\( V_k \) the variance of the traffic offered to link \( i \).
\( n_i \) the number of junctions required on \( i \) (regarded as a continuous variable for theoretical convenience.)
\( K_k \) \( \{ k \mid \text{link } i \in m^k \text{ for some } j \} \) \( j=1,\ldots,m_k \}
\( H_k \) \( \{ h^k \mid k \in K_k \} \) \( j=1,\ldots,m_k \}
\( I_k \) \( \{ i \mid \text{link } i \in m^k, (j=1,\ldots,m_k) \}

3. THREE CONCEPTS OF OPTIMALITY

Before proceeding to discuss the mathematical conditions of optimality for these three concepts it is necessary to introduce the concepts of marginal chain cost and chain cost per unit flow.

Definition 1 If \( C(h) \) is the total cost of the network, (as a function of chain flows \( h \)), then the marginal cost of chain \( m^k \) is defined as
\[
    D_k^h = \frac{\partial C(h)}{\partial h^k}
\]
Definition 2 For alternate routing networks, the cost per unit flow on chain \( m^k \) is
\[
    \bar{C}_k = \bar{a}_k \cdot \bar{C}_1
\]
where \( \bar{a}_k = \begin{cases} 
1 & \text{if link } i \text{ is on chain } m^k \\
0 & \text{otherwise.}
\end{cases} \)

The first concept of optimality is known as "System" optimality, and a network \( T \) is said to be "System optimized" if it is designed in such a way that the total network cost \( C(h) \) is a minimum, and the Origin-Destination demands have been met. A feasible chain flow pattern \( h(T) \) which minimizes the cost function \( C(h) \) is termed a System optimizing chain flow pattern.

Mathematically, a System optimizing flow pattern \( h(T) \) is a solution of the minimization problem:

Minimize \( C(h) \)
Subject to:
\[
    f_i = \sum_k a^k_i h_j^k \quad i = 1, \ldots, t
\]
\[
    h_j^k > 0 \quad (k = 1, \ldots, n)
\]
\[
    D^k_p \leq D^k \quad (1-B^k) t^k \quad k = 1, \ldots, n
\]

(Where \( B^k \) is a constant which specifies the Origin-Destination grade of service.)

By considering the Khun-Tucker [9] optimality conditions for the above mathematical program, the following property of a System optimizing chain flow pattern can be deduced.

Theorem 1 For a given network \( T \), a chain flow pattern is System optimal if for each OD pair \( k \), there is an ordering \( p_1, \ldots, p_s, p_{s+1}, \ldots, p_{m_k} \) of the chains joining \( O^k \) to \( D^k \) such that the following condition holds:
\[
    \lambda_k = D^k_{p_1} = \ldots = D^k_{p_s} < D^k_{p_{s+1}} < \ldots < D^k_{p_{m_k}}
\]
where
\[
    h_{p_r}^k > 0 \quad r = 1, \ldots, s
\]
and
\[
    h_{p_r}^k = 0 \quad r = s+1, \ldots, m_k
\]

The problem of finding the System optimizing flow pattern in a telephone network has been considered by Barry [1,2], and this will be discussed further in Sections 5 and 6.

The second concept, which has been considered by Sparrow and Dafermos [3,4] in relation to transportation networks, refers to the "User" optimizing flow pattern. This flow pattern has the equilibrium property that "no user has any incentive to make a unilateral decision to change his route." This pattern is expected to arise in networks where the individual users choose their routes independently. The equilibrium conditions are formally very similar to those of the System optimization; however, the chain costs per unit flow replace the role of the marginal chain costs.

Definition 3 A flow pattern \( h(T) \) is said to satisfy the equilibrium conditions for a User optimized network \( T \) if, for each OD pair \( k \), there is an ordering

\[
    c_1 = \begin{cases} 
    c_{1n} & \text{if } f_1 \neq 0 \\
    \lim_{f_1 \to 0} c_{1n} & \text{if } f_1 = 0
    \end{cases}
\]
of the chains joining $O_k$ to $D_k$ such that
\[
\mu_k = \rho_1 < \cdots < \rho_s < \cdots < \rho_{s+1} < \cdots < \rho_{n_k}
\]
where
\[
\rho_r > 0 \quad r = 1, \ldots, s
\]
\[
\rho_r = 0 \quad r = s+1, \ldots, n_k
\]
It can be shown ([5]) that for telephone networks the above conditions imply that chains with positive flow have the maximum traffic efficiency per unit circuit cost, and those chains with zero flow have a traffic efficiency per unit circuit cost which is less than (or equal to) the maximum.

The third principle of optimality arises when planners attempt to minimize the total cost for traffic flowing between a given Origin-Destination pair. In this case the optimizing principle can be seen as a game played by the Origin-Destination pairs which act independently of one another in order to minimize the cost of routing their traffic to the required destination. (The applicability of game theoretic principles to the OD pairs of a network was realized through discussions with Ristic [6]). For a given OD pair $k$, define the set $I_k$ as the set of links which form the permissible routes between $O_k$ and $D_k$ (see Section 2), then the total cost $C_k$ between $O_k$ and $D_k$ is
\[
C_k = \sum_{i \in I_k} c_i n_i
\]
Now, each OD pair $k$ adjusts the chain flows $h^k_j$ in such a way that $C_k$ is a minimum and the demand and non-negativity constraints are satisfied.

i.e. For each OD pair $k$ in the network
\[
h^k_j > 0 \quad j = 1, \ldots, n_k
\]
and $C(k)$ is a minimum.

The equilibrium conditions for the above problem can now be stated as follows:

Definition 4: A network $T$ is said to be "game theoretically optimized" if for each OD pair $k$ there is an ordering
\[
\rho_1, \ldots, \rho_s, \rho_{s+1}, \ldots, \rho_{n_k}
\]
of the chains joining $O_k$ to $D_k$ such that
\[
\frac{\partial C_k}{\partial \rho_r} = \cdots = \frac{\partial C_k}{\partial \rho_s} < \frac{\partial C_k}{\partial \rho_{s+1}} < \cdots < \frac{\partial C_k}{\partial \rho_{n_k}}
\]
where
\[
h^k_r > 0 \quad r = 1, \ldots, s
\]
\[
h^k_r = 0 \quad r = s+1, \ldots, n_k
\]

4. AN EXAMPLE.

In order to illustrate the three concepts of optimality discussed in the previous section; the following network (which has been adapted from Dafermos [4]) will be considered. The network is depicted in Fig.1 and it consists of two origin-destination pairs which use the five directed links labelled 1 to 5.

\[\text{FIGURE 1}\]

OD pair XY uses links 1, 3 and 5.
OD pair YX uses links 2 and 4.
The traffic demands for both OD pairs are 120 units.

\[
i.e. \sum_{j=1}^8 h^k_j = 120
\]

where the chain flows obviously corresponds to the link flows in this problem.

Hence
\[
h^k_1 = f_1; h^k_2 = f_2; h^k_3 = f_3
\]
\[
h^k_4 = f_4; h^k_5 = f_5
\]
The cost function is
\[
\sum_{i=1}^8 c_i n_i = C
\]
where the numbers of junctions are assumed to take the following form
\[
\begin{align*}
11 & = 4f_1 + 2f_2 + 900f_3 \\
12 & = 6f_1 + 2f_2 + 900f_3 \\
13 & = 6f_1 + 3f_2 + 820f_4 \\
14 & = 8f_1 + 3f_2 + 820f_4 \\
15 & = 5f_1 + 1220f_5
\end{align*}
\]
and the costs per junction for each link are $1.
Thus the marginal chain costs become
\[
\begin{align*}
16 & = 8f_1 + 4f_2 + 900 \\
17 & = 15f_1 + 6f_2 + 820 \\
18 & = 10f_1 + 1220 \\
19 & = 12f_1 + 4f_2 + 900 \\
20 & = 16f_1 + 6f_2 + 820
\end{align*}
\]
Using the above information, a nonlinear programming algorithm was used to locate the system optimizing flow pattern i.e. the flow pattern which minimizes the total network cost $C$. The details of the System optimizing flow pattern are summarized in Table 1.

The second principle of optimality (User optimization) can also be applied to this network, the chain cost per unit flow functions are:
\[
\begin{align*}
\text{Cl} & = 4f_1 + 2f_2 + 900 \\
\text{C2} & = 6f_1 + 3f_2 + 820 \\
\text{C3} & = 5f_1 + 1220 \\
\text{C4} & = 6f_1 + 2f_2 + 900 \\
\text{C5} & = 8f_1 + 3f_2 + 820
\end{align*}
\]
A modified nonlinear programming algorithm was applied and the User optimizing flow pattern was calculated. This flow pattern is summarised in Table 1.

The third principle of optimality involves the minimizing of the individual Origin-Destination costs, thus from the information given above it is clear that

\[
c_1 = 4f_1 + 2f_2 + 900r_1 + 6f_3 + 3f_4 + 820r_2 + 3r_3 + 1220r_4
\]

\[
c_2 = 6f_1 + 2f_2 + 900r_1 + 8f_3 + 3f_4 + 820r_2
\]

Differentiating the Origin-Destination costs with respect to the variables "controlled" by the Origin-Destination pairs, gives the following relations:

\[
\frac{\partial c_1}{\partial f_1} = 4 + \frac{\partial c_1}{\partial f_2} = 2 + \frac{\partial c_1}{\partial r_1} = 900 + \frac{\partial c_1}{\partial f_3} = 6 + \frac{\partial c_1}{\partial f_4} = 3 + \frac{\partial c_1}{\partial r_2} = 820 + \frac{\partial c_1}{\partial r_3} = 3 + \frac{\partial c_1}{\partial r_4} = 1220
\]

Application of the modified nonlinear programming algorithm enabled the calculation of the chain flow pattern corresponding to the third principle of optimality and this pattern is also given in Table 1.

### Table 1: Optimal solutions to the Three Concepts

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>USER</th>
<th>GAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1</td>
<td>50.2</td>
<td>61.3</td>
</tr>
<tr>
<td>f2</td>
<td>66.1</td>
<td>64.3</td>
</tr>
<tr>
<td>f3</td>
<td>35.2</td>
<td>47.9</td>
</tr>
<tr>
<td>f4</td>
<td>53.9</td>
<td>55.7</td>
</tr>
<tr>
<td>Total</td>
<td>317,557</td>
<td>321,938</td>
</tr>
</tbody>
</table>

Table 1 clearly shows that the three principles of optimality, applied to this simple network, have produced three different "optimal" chain flow patterns, and they each have a different total network cost. Furthermore, Table 2 summarizes the Origin-Destination costs for the three optimizing principles, and it demonstrates that minimizing the individual Origin-Destination costs (the third principle) is not equivalent to minimizing the total network cost.

It is interesting to note that for OD pair X-Y, the game theoretic approach has located a cheaper origin-destination cost than the equivalent OD cost in the System optimized solution, however, the "selfishness" of the XY pair in the game optimal solution. This result is not unexpected because it is a well-known property of games that the optimal set of strategies for a non-cooperative game, is not optimal (in general) for the game in which the players cooperate to achieve an overall minimum cost (System optimization).

### Table 2: Origin-Destination Costs for Optimal Solutions

<table>
<thead>
<tr>
<th>X-Y OD Cost</th>
<th>Y-X OD Cost</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>152,085</td>
<td>165,472</td>
<td>317,557</td>
</tr>
<tr>
<td>152,086</td>
<td>166,919</td>
<td>321,938</td>
</tr>
<tr>
<td>151,460</td>
<td>166,919</td>
<td>318,279</td>
</tr>
</tbody>
</table>
assumed that the rest of the network has already been designed and a small portion of the network is isolated for a particular OD pair. "Optimizing" equations for this OD pair are derived subject to the last choice link grade of service criterion.

(11) Application of the optimizing principles to the complete network. In this stage Pratt attempts to relate OD cost minimizations to the optimization of the complete network, while Wallstrom describes an iterative procedure which is expected to produce the optimal design for the entire network.

It is important to note that there is no mention in Pratt's or Wallstrom's articles that the above minimization is subject to the non-negativity requirements for numbers of junctions, i.e. $n_i > 0 \; \text{for} \; i = 1, \ldots , l$.

If this requirement is included in the mathematical analysis of the model, it becomes clear that the optimality conditions are incomplete, and Khun-Tucker theory [9] must be used to obtain the conditions for a stationary point, i.e. $\frac{\partial f_i}{\partial n_i} |_{n_i} = 0$.

The conditions (21) mean that it may be possible to reduce the cost further by decreasing $n_i$, except that this would result in the provision of negative numbers of junctions on link $i$. In section 10 of his paper, Pratt points out that solution of the optimizing equations may lead to a value for the marginal occupancy which is greater than unity and this is recognized as a condition for rejecting a high usage link as being uneconomic for $0_k$ to $D_k$ traffic. It is interesting to note that no mathematical criterion for rejecting $0_k$ to $D_k$ traffic from parts of the alternate route has been devised. In general, rejection of an alternate route has been devised during the course of manual computations when it is realized that the route is very expensive by comparison with other choices.

6. DISCUSSION

In section 4 it was demonstrated that even with particularly simple networks the three principles of System, Game and User optimization do not coincide. However, when the principles of optimization were applied to Berry's model, a number of interesting results were obtained. Firstly, when a small telephone network with four exchanges and two tandems was System, Game and User optimized using a modified nonlinear programming algorithm, it was found that all three optimizing flow patterns were identical, i.e. System, Game and User optimizing the network were equivalent procedures for this particular problem. The reasons behind this result are thought to be due to the very high costs of the links in the alternate routes of the network, since it was found that only the direct routes needed to be used to obtain a minimum cost network. When the principles were applied to a larger and more realistic network (such as the Adelaide Telephone Network) it was found that the System and Game optimized chain flow patterns were identical, but the User optimized chain flow pattern was different and it resulted in a more expensive value for the cost function. On the basis of this result it is natural to ask why User optimization is studied in relation to telephonc networks. The answer to this question is related to the relative speeds of computing the System and User optimizing flow patterns. Experience has shown that User optimization is simple to program for a computer and is very rapid. Also, a number of examples which have been User optimized and System optimized have shown that the difference in total network costs between these two flow patterns is usually small in relation to the total network cost of the System optimized solution. For this reason it is hoped that a System optimized solution can be reached more rapidly by User optimizing the network first, and then iterating from the User solution down to the System solution.

A second result of applying the optimizing principles to Berry's model concerns the equivalence of the System and Game optimized solutions. The following theorem demonstrates the equivalence of these two optimizing principles when applied to Berry's model.

Theorem 2 If $K_1$ is the set of OD pairs $k$ such that link 1 is on some chain $m_k$ of OD pair $k$, and $K_2$ is the set of all chain flows $h^k$ such that $k$ is in $K_1$, and the link costs $c_{i1} n_i$ in a network are only functions of the chain flows in $h^k$ (for each link 1), then System optimization is equivalent to minimizing the total OD costs $C^k$ for each OD pair $k$.

Proof: $C^k = \sum c_{i1} n_i \hspace{1cm} \text{for } i \in I_k$

Hence $\frac{\partial C^k}{\partial n_i} = \sum c_{i1} \frac{\partial n_i(h_i)}{\partial n_i} |_{n_i} = 0 \; (j=1, \ldots , m_k)$

But $C = \sum c_{i1} n_i |_{i \notin I_k} + \sum c_{i1} n_i \hspace{1cm} \text{for } i \notin I_k$

Total Network Cost

Thus $\frac{\partial C}{\partial n_i} = \frac{\partial C^k}{\partial n_i} \hspace{1cm} \text{for all OD pairs } k$ in the network and hence conditions (8) are exactly equivalent to conditions (12).

Q.E.D.

The reason for this theorem being applicable to Berry's model is that the chain flows originating from different OD pairs are assumed to be statistically independent and that the link costs $(c_{i1} n_i)$ are only functions of the chain flows in the set $q_i$.

It is difficult to justify this approach for the following reason: if two origin-destination pairs $k_1$ and $k_2$ share a common link 1 (high usage), then two OD cost functions will be functions of $n_1$ and this leads to three optimizing equations:

$$\begin{align*}
3c_{11} \hspace{1cm} 3c_{21} = 0 \hspace{1cm} 3c_{11} = 0 \hspace{1cm} \text{and} \hspace{1cm} 3c_{21} = 0
\end{align*}$$

and it is not clear that one value for $n_1$ will minimize the two Origin-Destination costs and the total network cost simultaneously.

From remarks made in Wallstrom's article it appears that in order to minimize the total network cost it is only possible to minimize the total network cost if the OD costs are independent.
necessary to apply the optimizing equations obtained for each OD pair in an iterative manner, proceeding from lower to higher levels in the network and designing the final choice routes from the given grade of service requirements. In proposing this iterative scheme for optimizing a complete network, Wallstrom must also justify minimizing the OD costs with respect to the numbers of junctions. It should also be clear from earlier remarks (section 5), that the equations (22) are incomplete, as each minimization is subject to the requirement that the numbers of junctions are non-negative.

It appears from the foregoing analysis that the principles of minimizing the total network cost and minimizing the OD costs with respect to the numbers of junctions on high usage routes have been "combined" to give a network which is not necessarily optimal. In fact, Rapp was aware [10] in 1964 that the Constant Background Traffic Model when applied to a certain small triangular network gave an erroneous "optimal" solution. The true optimal network design had no junctions provided on the alternate route whereas the Constant Background Traffic model gave a network design in which the expensive alternate route was required. It would seem that the difficulties encountered by the Constant Background models are partially due to the incompleteness of the optimality conditions, and also to invalid combination of network optimizing principles.

7. CONCLUSIONS

The question of providing sufficient junctions between exchanges for anticipated demand is of vital importance to Teletraffic Planning authorities. These planning authorities hold a unique position, in that, unlike road traffic planners, they are able to specify the routes to be taken by traffic with the aid of exchange switching equipment. For road traffic it is usually necessary to regard the users of the network as acting independently of one another, and hence it is to be expected that these users will try to minimize their Origin-Destination costs. However, no such assumption needs to be made for telephone subscribers as the equipment selects their routes; but this fact should not prevent investigation of road traffic optimizing principles (such as "User"), because it may be possible to achieve more rapid methods of optimization for telephone networks.

This paper has attempted to demonstrate that a closer examination should be made of the optimizing principles which are currently used by many teletraffic planning authorities and mathematical justification for the assumptions made in these models should be investigated in the light of the remarks made in sections 5 and 6 concerning the Constant Background Traffic Models.

8. ACKNOWLEDGMENTS

The author wishes to express his appreciation of the guidance and encouragement given by Professor R.B. Potts and Dr. L.T.M. Berry of the Department of Applied Mathematics, University of Adelaide. Thanks are also due to Dr. C.W. Pratt and Mr. K. Vawser of the Australian Post Office for their many useful discussions. Work in this project is supported by a contract with the Australian Post Office.

9. REFERENCES

7. C.W. Pratt "The Concept of Marginal Overflow in Alternate Routing", Australian Telecommunication Research, 1967, 1, [1,2], pp 76-82.