SIMPLIFIED PROCEDURES FOR NETWORK PRICING

George Yarrow

Research Paper Series

Regulatory Policy Institute

Oxford, May 1992

SIMPLIFIED PROCEDURES FOR NETWORK PRICING *

George Yarrow

Hertford College, Oxford OX1 3BW

* The contents of this paper are preliminary. Comments are very welcome, but material should not be quoted without the permission of the author.

SUMMARY

This paper sets out results from work in progress on the issue of tariff structures for access to, and use of, network facilities. Starting from a simple version of the notional path approach, it is argued that it is relatively straightforward to extend the method so as to incorporate genuine network features into the resulting tariff structures. The resulting approach is called least-cost notional flow analysis. The generalisation leads naturally to a tariff structure based on entry and exit charges. It also offers a framework in which conflicting pricing objectives -- such as cost reflectivity, cost recovery, spatial averaging of prices, ease of monitoring, etc. -- can be traded-off in a relatively transparent way.

1. Introduction

The increased emphasis that has recently been placed on the goal of developing competition in industries such as electricity and gas has generated growing interest in the important issue of network pricing. Given naturally monopolistic cost conditions in the transmission and distribution activities of these industries, the development of effective competition in supply depends heavily upon the establishment of conditions under which a number of different competing firms can make use of a common set of network facilities such as wires or pipelines. Further, the terms on which access to, and use of, the common transmission and distribution facilities is made available to different firms will have major implications for the nature of competition in both the wholesale and retail markets. For example, where the naturally monopolistic network operator is also one of the competitors in downstream retail supply, there is the potential for transmission and distribution charges to be used strategically to place downstream rivals of the network operator at a competitive disadvantage, thus distorting competition in supply.

Even where there is clear business separation between transmission or distribution activities and wholesale or retail supply -- so that, for example, network pricing decisions are made on the basis of their contributions to the profits of the transmission or distribution businesses alone (and not to profits from supply businesses as well) --access or use-of-system charges can have substantial implications for competitive conditions elsewhere in the industry. Being "transportation" activities, the transmission and distribution costs incurred by the network operator are heavily influenced by geographical factors. Hence the precise ways in which the network operator's own costs are, or are not, reflected in network prices will have major implications for the degree of spatial differentiation in the final supply prices of the commodities being transported.

To illustrate, where there exist incumbent suppliers with some element of fixity in the locations of their inputs and outputs -- as in electricity generation, for example -- the relative costs of suppliers (whether incumbent or entrant, wholesaler or retailer) can be significantly affected by changes in the structure of transmission and distribution charges. This is most obvious in the case of an input source that, compared with other sources, is relatively distant from the main centres of demand. In such a case a move to increase distance-related charges for use of the network will tend to place the relevant supplier at a cost disadvantage compared with rivals. More generally, the spatial structure of network prices will have very important effects on the degree of spatial segmentation in the markets for wholesale and retail supplies, and spatial segmentation will in turn have significant implications for competitive conditions in those markets.

Formal economic principles for network pricing have been most comprehensively and elaborately developed for electricity networks by Bohn, Caramanis, Schweppe and Tabors[1]. These models take specific account of the main physical principles governing flows in electricity networks and derive expressions for spot prices of electricity differentiated by both time and location. Similar procedures can also be used in the case of gas networks, but taking account of the somewhat different physical relationships governing flows of gas. In both cases the end result is a set of expressions which reflect the marginal economic costs incurred as a result of changes in inputs to and outputs from the network at different times and at different places.[2]

In contrast to the apparent sophistication of this underlying theory, practical implementations of network pricing have to date tended to be both few and simple, in large part because the pricing problems have traditionally been sidestepped by policy approaches based upon vertical integration between "transport" and "supply" activities within the context of franchised monopoly. The most common starting point for practical implementations of network pricing has been the identification of a "contract" or "notional" path, along which the commodity is deemed to be transported by the network operator and which is therefore employed as the basis for charging the user of the network. It is, however, well known that such notional paths may bear little relationship to the actual physical flows in networks, and hence that the resulting charges for use of the network may bear little relationship to the economic costs imposed (either on the network operator or on others) by any particular user.

The gap between spot pricing based on marginal costs and the implementations of network pricing based on notional paths is wide. In section 2 below, I will argue that this is not entirely a case of practice lagging behind theory in that there are good reasons why a "strict" marginal-cost approach is not necessarily a desirable aim of regulatory policy. Nevertheless, there does appear to be considerable scope for developing network pricing structures in ways that both improve their ability to signal underlying economic costs and, simultaneously, satisfy criteria such as simplicity and transparency that may be important within a regulatory context. Sections 3 to 8 below seek to do precisely that, by

---

1 R.E. Bohn, M.C. Caramanis and F.C. Schweppe, "Optimal pricing in electricity networks over space and time", Rand Journal of Economics, Autumn 1984. F.C. Schweppe, M.C. Caramanis, R.D. Tabors, and R.E. Bohn, Spot pricing of electricity, Kluwer Academic Publishers, Norwell MA, 1988.

2 The emphasis on economic costs here serves as a reminder that the relevant costs may not be borne by the network operator. For example, charges may reflect the costs imposed by one network user on another as a result of additional congestion on the system.

outlining one means of developing notional path analysis into something that more closely approximates actual flows in a network. The approach is labelled least-cost notional flow analysis. The procedures outlined have the advantage that, within a well-defined framework, the degree of approximation to underlying economic costs is partly a matter of choice. In terms of regulatory practice this means that tariff structures can be modified gradually without major discontinuities in pricing philosophy.

The analysis is organised as follows. Section 3 outlines the simplest type of linear network and shows why notional path analysis can lead to tariff structures that bear very little relationship to transmission and distribution costs. It is, however, also shown how, in this example, implicit exchange or "swapping" of commodity does in fact lead to a structure of payments for use of system that reflects the network operator's own costs. In section 4, it is shown how the approach can be formulated as a linear programming problem. This leads in a straightforward way to a generalisation to complex networks which is outlined in section 5. Section 6 provides worked examples for both nodal and zonal implementations of the approach, while section 7 discusses a variety of possible refinements, including peak-load pricing, non-linear pricing, and the incorporation of non-spatial cost drivers (eg. step changes in voltage or pressure levels). Section 8 considers some additional issues concerning alternative methods of recovering full costs, including the question of the appropriate balance between capacity and commodity charges. Finally, section 9 summarises the main conclusions.

2. The principles of network pricing

The theoretical development of expressions for spot-prices in transmission and distribution networks has taken place in what might loosely be termed a "traditional" public utility context. That is, the work can be seen as an extension of the type of marginal-cost approaches pioneered by Boiteaux, Turvey, and others. The central concern of this work is with the efficient allocation of resources within a static context free of fundamental information problems.[3] As noted in the introduction, however, much of the current interest in network pricing stems from public policy initiatives to facilitate and promote the development of competition in supply businesses (which rely upon naturally monopolistic transmission and distribution inputs but which are not themselves naturally monopolistic). This necessarily leads to a somewhat different perspective on the issues.

[3] The high-water mark of this approach in the UK was the 1967 White Paper on Nationalised Industries.

The two approaches -- promotion of allocative efficiency and promotion of competition -- are not, of course, mutually exclusive, and regulators may be concerned both to promote efficient signalling of relevant network costs and to encourage competition in related activities. There will, however, be trade-offs between the two approaches, and it is important to recognise these at the outset. Indeed, at least in respect of the position in the UK, it can be argued that the development of the policy objective of promoting competition was itself partly a response to the weaknesses of the traditional, "allocative efficiency" approach to public utility pricing.

Consider, for example, the pitfalls of the marginal cost pricing philosophy. First, in more or less any real, practical example, it is simply wrong in theory to hold that marginal-cost pricing leads to allocative efficiency. As a consequence of the ubiquity of imperfectly competitive markets,[4] the achievement of such efficiency, even in a static context, requires that a whole raft of second-best factors be taken into account. These include factors such as price-elasticities of demand, cross-price elasticities of demand, and the price-cost margins of substitute or complementary products. When translated into the language of business, these factors become familiar considerations like "what the market will bear" and "our competitors' pricing policies". Information on the cost of providing the relevant good or service remains an important input into efficient pricing decisions, but it is only one of several such inputs.

Second, as already indicated, the relevant public policy objective may not simply be the promotion of allocative efficiency. Distributional considerations typically play a large role in regulatory policy and, in the context of network industries, spatial distribution is a particularly important consideration. Internationally and historically, regulatory practice reveals a strong underlying "preference" for pricing structures that, if not completely then at least to a significant degree, suppress major spatial variations in prices.

Third, and perhaps most fundamentally, information on marginal costs is itself economically costly to obtain, implying that account needs to be taken of problems of imperfect and asymmetric information. In consequence, and as recognised by critics of the traditional public utility approach to pricing well before the more recent explosion of theoretical interest in informational problems, estimates of marginal costs are necessarily

4 Standard economic terminology is unfortunate here in that it tends to convey the impression that imperfect competition is in some sense "flawed" competition. In fact, some or other form of imperfect competition is almost always to be preferred to perfect competition (which generally tends to be infeasible anyhow).

subjective.[5] Monitoring of performance (by regulators, say) in this area is therefore particularly problematic.

It is from the third of the above critiques that the more recent stress on the objective of competition most naturally flows. For, among other things, competition is a means by which incentives to discover information are firmly established, a point consistently stressed in the work of Friedrich Hayek. Put very loosely, the shift in public policy reflects the replacement of nominal[6] preoccupation with static efficiency with a much greater concern for dynamic efficiency.

Once the focus is changed from considerations of static allocative efficiency to considerations of competition, there can in fact be an efficiency case, as well as a distributional case, in favour of tariff structures that suppress spatial differentiation of prices, at least to some degree. To users of the network, charges for transmission and distribution are simply forms of transport cost, and, where they are economically significant, transport costs tend to lead to geographic segmentation of markets.

Geographic averaging of transmission and distribution charges -- that is making the charges relatively insensitive to the particular locational characteristics of individual transport requirements -- will, then, have the effect of reducing barriers to trade, thereby helping to create a larger, more integrated market capable of efficiently sustaining a larger number of competitive suppliers.

Striking an appropriate balance between the efficiency benefits of (a) providing networks users with accurate signals of the costs their activities impose and (b) promoting competition is likely to be a highly inexact exercise, largely because of the difficulties of quantifying the various effects (particularly over the longer term). This is not, however, the only issue to be resolved in developing tariff structures. Also important, given the regulatory context, are factors such as transparency and ease of monitoring of the charging system.

---

[5] See J. Wiseman, "The theory of public utility price -- an empty box", Oxford Economic Papers, 1957, and G.F. Thirlby, "Economists' cost rules and equilibrium theory", Economica, 1960.

[6] It is doubtful that promotion of allocative efficiency has generally been the major, real objective of those responsible for the regulation of network industries. Otherwise, and given the extent of spatial variations in costs, it is difficult to account for the lack of spatial variation in prices.

These latter factors place a premium on objectivity in tariff construction, which in regulatory practice has frequently led to pricing structures based upon accounting methods such as fully-allocated costs. However, the objectivity of such methods can, on closer inspection, turn out to be something of an illusion, involving, for example, highly arbitrary cost allocations. What appears to be most important for transparency and monitoring, therefore, is not so much the use of accounting information *per se* but rather the derivation of charges from that information according to well defined procedures.

The use of procedural rules (or "rules of thumb") in pricing is well understood from empirical studies of pricing behaviour in reasonably competitive markets. Rules of thumb can be efficient mechanisms for decision making under conditions of uncertainty when the evaluation of the likely consequences of actions is expensive or infeasible. In a regulatory context, procedural rules have additional benefits both to the regulator -- in terms of monitoring -- and to third parties affected by network charges. Where, for example, the tariff structure is based upon the application of relatively simple and agreed transformations of accounting data, there is at least some prospect of affected parties being able to make meaningful forecasts of future charges.

3. Notional or contract path charging systems

As noted in the introduction, tariff structures based upon (a) the identification of a notional or contract path for a particular transport requirement and (b) the setting of charges according to some allocation of accounting costs to identified paths have, hitherto, been the most commonly found approach to network pricing, and their popularity has not been wholly unjustified. The advantage of such an approach lies in its simplicity and ease of application, which gives it transparency and monitoring advantages over the much more sophisticated, "strict" marginal cost methods that have been proposed.

Notional or contract path charging systems do, however, have major drawbacks associated with their frequent inability to transmit appropriate cost signals to users of the network. The reasons for this can be illustrated by reference to the simplest type of linear network.

Figure 1 shows two locations or areas, labelled A and B, which are connected by a single transmission line of unit length. For simplicity, assume that (a) the capital costs of constructing the link are £c per unit of flow per unit of distance and (b) the variable costs of transporting the commodity are zero.[7] Now suppose that a supplier with inputs at location A wishes to supply 100 units to a customer at B, while a supplier at B wishes to supply 50 units to a customer at A. Under the contract path approach, both supply patterns would be deemed to have made use of the same length of transmission capacity and, in the simplest case, both supplies would be charged the same rate per unit of commodity transported. The total charge for the A to B supply would therefore be twice the total charge for the B to A supply, since the former flow is twice the magnitude of the latter.

Figure 1. A point-to-point transmission link.

A		B
S: 100		S: 50
D: 50		D: 100

However, these charges bear little relationship to the costs borne by the network operator in meeting the transportation requirements of its two customers. On the assumptions made, the actual amount of the commodity transported will be 50 units, in a direction from A to B. The transmission capacity required to meet the overall supply/demand pattern is therefore 50 units, so that the total cost of the link is just £50c. If this is allocated between the two supplies in proportion to load size (and on a simple contract path basis) then the price per unit transmitted is just £c/3.

Consider now the impact on the network operator's costs of an increase in one unit in the supply from A to B. This will require an increase in the capacity of the link by one unit, imposing an incremental cost of £c. The charge for supply from A to B is therefore equal to one third of the incremental cost. On the other hand, if there is an increase in one unit in the supply from B to A this will reduce the required capacity of the line by one unit, since now only 49 units per period will flow from A to B. In this case the incremental cost imposed by the increase in demand is negative, and equal to -£c. Thus, if the notional path charges of the previous are used, the A to B shipper pays substantially less

7 The assumption of zero variable costs is a good first approximation for a gas pipeline operating under free flow. It is much less appropriate for an electricity link, where energy losses are likely to be a significant factor.

than incremental cost while the B to A shipper pays substantially more than incremental cost. The result will therefore tend to be an inefficient pattern of location of commodity supplies and demands as users of the network respond to the price signals.

The problem with the simple notional or contract path approach in this example is that it takes no account of the possibility of re-allocating supplies to demands so as to reduce transport requirements. Thus, in effect, the demand for 50 units at A is met from supplies at A, not from B as would be indicated in the relevant supply contract. In an industry such as electricity, this re-allocation occurs automatically as energy flows respond to electrical pressures. However, similar re-allocations can also take place in other circumstances where definite actions are required to achieve the desired effect.

A good example is petrol swaps, whereby competing suppliers of petrol may exchange product in order to reduce transport costs. Thus, if company 1 has a refinery at A, company 2 has a refinery at B, and each company wishes to supply customers in both areas, then transport costs are reduced if company 1 swaps petrol from its refinery with petrol from company 2's refinery. In this way each company can draw at least part of its requirements for supplies to its own "distant" location from a more favourably located production facility, thus reducing "cross-hauling". The practice is discussed in the Monopolies and Mergers Commission Report on the Supply of Petrol (1990).

In the linear network example, a cost-reflective tariff structure can emerge if each customer is charged £c per unit but commodity swaps (either explicit or implicit) are allowed. Without swaps payments are £100c for A to B transport and £50c for B to A transport. On the other hand, a swap between suppliers of 50 units of input at A for 50 units at B will reduce the total payment to £50c, which then has to be shared between the two users. If the allocation is such as to correspond to the marginal contributions to post-swap payments of the two different transport requirements -- and this is what would be expected in competitive commodity markets -- the resulting payments incurred by network users will be £c per unit for A to B transport and -£c per unit for the reverse direction, which correspond with the costs incurred at the margin by the network operator.

Swapping of supplies to reduce costs relies upon homogeneity of the underlying product. It can be noted, however, that, even if the commodity being transported from A to B is different from that being transported from B to A -- so that swaps are inappropriate -- the simple expedient of charging the same rates for transport in both directions is likely to produce poor cost signals to users.

Suppose, for example, that the link between A and B represents a road, rail or shipping route, and suppose further that the demand for transportation of goods from A to B is substantially above that for transportation from B to A. Then the transport company may well charge a back-haul rate, offering lower prices for shipments from B to A than for shipments from A to B. This is similar to a peak-load pricing scheme, in that, since the A to B demand determines capacity requirements, capacity charges are allocated to that particular direction. The underlying point is that a particular input or combination of inputs (road and lorry, rail line and train, harbour and ship) leads to the joint supply of two distinct outputs (A to B transport and B to A transport).[8,9]

4. Least-cost notional flow analysis

Because of the simplicity of the linear network problem described in section 3, it was very straightforward to arrive at the incremental costs imposed on the network operator by the hypothesised changes in the commodity supply/demand pattern. In this section I will outline an alternative method or procedure for getting to the same tariff structure. This formalises the swapping arrangements discussed above.

Table 1. Contract-cost matrix for linear network

	A	B
A	0	c
B	c	0

[8] It can be noted that the post-privatization tariff structure for third-party use of British Gas's pipeline network contained "back-haul" discounts, which modified the notional path charges on some routes.

[9] Note that one difference between the two types of network is that marginal costs on "back-haul routes" will tend to be negative for homogeneous commodities (eg. electricity and gas) and positive, albeit relatively low, for differentiated commodities (eg. telecoms, rail and postal services).

Suppose that each of the two suppliers has a contract with the network operator for transport of the commodity. The contracts are differentiated by the direction of flow, but the charge per unit is £c per kilometre in both cases, being derived from the network operator's incremental cost of upgrading the transmission link by one unit. Then the contract costs to system users of having the network operator transport the commodity can be represented by the matrix shown in Table 1, where rows show sources of supplies and columns show destinations.

Now suppose that notional flows are allocated so as to minimise the total contractual payments due to the network operator (i.e. to minimise the costs incurred by network users in obtaining the requisite transportation services). On the assumptions made thus far, and given the magnitudes of the supplies and demands for the commodity at each location, this defines a linear programming problem to be solved. Letting

x_{aa} be the notional "flow" of the commodity from A to A,

x_{ab} be the notional flow of the commodity from A to B,

x_{ba} be the notional flow of the commodity from B to A, and

x_{bb} be the notional "flow" of the commodity from B to B,

the problem is:

$$\text{minimise } cx_{ab} + cx_{ba}$$

subject to

$$\begin{aligned} x_{aa} + x_{ba} &\geq 50 \\ x_{ab} + x_{bb} &\geq 100 \\ x_{aa} + x_{ab} &\leq 100 \\ x_{ba} + x_{bb} &\leq 50 \\ x_{aa} &\geq 0 \\ x_{ab} &\geq 0 \\ x_{ba} &\geq 0 \\ x_{bb} &\geq 0 \end{aligned}$$

The first two constraint inequalities here state simply that the amount of commodity transported to a particular location, including that allocated from supplies already available at the location, must be at least equal to the demand requirements at that point. The second pair of inequalities state that the amount of commodity transported from a particular location, including that allocated to meeting demand at the same location, must not exceed the amount available at that point.

The solution of the above problem is relatively trivial and is shown in the matrix of notional flows in Table 2. Thus 50 units of supply of the commodity at A are allocated to demand at A and 50 units are allocated to demand at B, while all the available supply at B (50 units) is allocated to demand at B.

Table 2. Least-cost allocation of notional flows

	A	B
A	50	50
B	0	50

The total, minimised, contractual payments are £50c, which, in this example, is the total capital cost actually faced by the network operator. Similarly, the notional flow on the network is, in this case, identical to the real flow. These equivalences arise because contract prices perfectly reflect actual costs. In more complicated cases, where charges are not perfectly aligned with the network operator's own costs, these equivalences will not generally hold. In particular, the notional flow allocations in the network will not correspond exactly with real physical flows.

The next step is to move from the notional flow allocation that minimises total contract costs to a set of prices that can be charged to system users. This is straightforward in that the solution to the linear programming problem automatically generates a set of "shadow prices" (which are technically the solution to the dual programming problem). Since the problem specified is a transportation problem, the shadow prices take a specific form: there is one price for each supply source and one for each demand sink. That is, the implicit tariff structure takes the form of a series of entry charges, levied on inputs to the network, and exit charges, levied on abstractions from the network.[10]

Linear programming theory further indicates that there is one degree of freedom in calculating the entry and exit charges associated with a particular, optimised, notional allocation of flows in the network. Intuitively, this occurs because the charge for the transport of some given quantity of the commodity will be the sum of an entry charge and

[10] See H.A. Taha, Operations Research: An Introduction, Collier Macmillan, for further details of the various results of linear programming theory cited in this paper.

exit charge. Only one "transportation service" is supplied and charged for, namely transport from A to B. All that matters, therefore, is that the sum of the relevant entry and exit charges equals the contract cost of that service. If the latter amounts to £p, say, then entry and exit charges of £1 and £(p-1) would produce the right answer; but so would charges of £2 and £(p-2), of £3 and £(p-3), and so on. We therefore need (arbitrarily) to fix one of the entry/exit charges before all the others can be calculated.

If we attach entry charges E_a and E_b to the rows of the optimised, notional flow matrix and exit (or "demand") charges D_a and D_b to the columns, linear programming theory tells us that, for any non-zero cell of the matrix, the corresponding entry and exit charges should sum to the cost shown in the associated contract cost matrix (i.e. the matrix shown in figure (i)). The equations for calculating entry and exit charges in the simple linear example are therefore as follows:

$$E_a + D_a = 0$$
$$E_a + D_b = c$$
$$E_b + D_b = 0.$$

By virtue of the fact that the transport charge for delivery of the commodity at any location from inputs at that same location is zero, it can be seen that the entry charge at each location is simply the negative of the exit charge. As will be seen later, this relationship between entry and exit charges does not hold in all cases that are of interest.

The "symmetric" solution to the above equations is:

$$E_a = D_b = c/2$$
$$E_b = D_a = -c/2,$$

although the following would do equally well (for any value of y):

$$E_a = c/2 + y$$
$$E_b = -c/2 + y$$
$$D_a = -c/2 - y$$
$$D_b = c/2 - y.$$

Assuming that these prices are charged, it can quickly be verified that the following charges per unit would be levied on system users:

For transport from A to A or B to B	0
For transport from A to B ($E_a + D_b$)	c
For transport from B to A ($E_b + D_a$)	- c

The charges therefore correspond to the underlying capital costs incurred by the network operator.

5. Generalisation of the least-cost notional flow approach

The generalisation of the above approach is conceptually straightforward. Let the network have N nodes, such that available supply at node i is S_i and demand at node i is D_i. Let the contract cost (i.e. charge to the user) for transporting the commodity from node i to node j be p_{ij}. Then the problem is to allocate notional flows through the network in a way that minimises total contract costs (i.e. total costs incurred by network users, not by the network operator). Let the notional flow between nodes i and j be x_{ij}. Then the formal programming problem is as follows:

$$\textit{minimise} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} x_{ij}$$

subject to

In linear programming terms, this is a trans-shipment problem.[11]

$$\sum_{j=1}^{n} x_{ij} \leq S_i \quad \textit{for all} \;\; i,$$

$$\sum_{i=1}^{n} x_{ij} \geq D_j \quad \textit{for all} \;\; j,$$

$$x_{ij} \geq 0 \quad \textit{for all} \;\; i, j.$$

---

[11] That is, it is a category of transportation problem in which sources of supply can also be destinations for deliveries (as they obviously can be at a node of an electricity network, say). See Taha, op cit, for more detail.

In building up the contract cost matrix, it is not necessary to specify a charge for every pair of points in the network. An alternative approach is to specify charges only for those points that are physically connected and adjacent to one another. Suppose, for example, that node i is connected to node j which in turn is connected to node k, so that transport from i to k involves trans-shipment through node j. Then, provided that p_{ij} and p_{jk} are specified, there is no need to specify p_{ik}. In complex networks this has the advantage that, if the charges for transport between adjacent, connected points are set to reflect the network operator's costs, then the charge for transport between more distant points, which will be derived from the optimisation problem, will also automatically be cost-related. There is, therefore, less scope for the emergence of anomalies in the tariff structure.

Associated with the allocation-of-flows problem is a dual programme, whose solution yields a set of dual prices (u_i,v_i) with the property that:

$$u_i + v_i = p_{ii}, \text{ for all i, and}$$

$$u_i + v_j = p_{ij}, \text{ where } i \neq j,$$

for any non-zero cell (i,j) in the matrix of notional flows that minimise contract costs incurred by network users. These non-zero cells represent connections along which notional flows are allocated in the solution of the minimisation problem. In general there will be n-1 of them, which, together with the n equations associated with the cells along the main diagonal of the contract cost matrix, give 2n-1 equations in 2n unknowns (the shadow prices). As stated earlier, therefore, there is one degree of freedom in the shadow prices.

The interpretation of the shadow prices or dual variables runs along standard lines. Consider the effect of a one unit increase in supply at node i and a one unit increase in demand at node j. Solving the problem of allocating nominal flows both before and after this change in the commodity supply/demand pattern, we will have two (optimised) levels of total contract costs. By standard theorems of linear programming it can be shown that:

$$u_i + v_j = \delta TP_{ij}$$

where δTP_{ij} is the change in minimised contract costs resulting from the change in the commodity supply/demand pattern.

It can be seen, therefore, that the sum of the shadow prices is equal to the change in the total, minimised charges to network users that occurs as a result of the additional supply and demand. Put another way, the sum of the two shadow prices is the marginal contract cost of the demand for transportation from node i to node j.[12] This indicates that u_i is a shadow price associated with inputs of commodity into the network, whereas v_j is a shadow price associated with abstractions of commodity from the network. The problem therefore leads naturally to a tariff structure based upon sets of entry charges (u_i) levied on commodity supplies into the network and exit charges (v_j) levied on commodity abstractions (demands) from the network.

Given the structure of the problem and given an initial solution, it is clear that an incremental demand for transport from i to j can (notionally) be met by assigning an extra unit of flow to the least cost direct route from i to j (that is the route that would be chosen in the notional flow problem if there were only the one transportation demand on the system). Let the resulting increment in contract costs be p^*_{ij}.[13] Then the optimum solution to the problem of allocating notional flows must give rise to a cost change that is no greater than p^*_{ij}:

$$\delta TP_{ij} \leq p^*_{ij}.$$

Now p^*_{ij} can be interpreted as the charge that would be levied on the basis of a simple notional path approach that paid no regard to network characteristics. Compared with this, the least-cost notional flow approach (which does incorporate network characteristics) can be interpreted as giving rise to a "network rebate" of $p^*_{ij} - \delta TP_{ij}$ to users of the system who have transportation requirements that give rise to (contract) cost-reducing reallocations of other notional flows. In respect of the simple linear network examined in sections 3 and 4, for example, the "network" rebate for A to B flows is zero (extra transportation in this direction does not affect previous allocations of nominal flows) and the rebate for B to A flows is -£2c. The explanation of the latter is that an increment in demand for transport from B to A is credited with the benefits of a reduction in the nominal flow allocated in the A to B direction.

---

12 This helps again to explain why there is one degree of freedom in the shadow prices. Two prices are linked to one incremental cost.

13 Where i and j are adjacent, connected nodes, then $p^*_{ij} = p_{ij}$.

6. Examples of the least-cost notional flow approach

6.1 A five-node case

To illustrate some of the above ideas, consider the five-node network illustrated in Figure 2, with supplies and demands at each node as specified in Table 3 and unit transport charges as shown in the matrix in Table 4. Such charges may, for example, be proportional to the lengths of wires or pipes connecting the relevant points. Note that costs are included only for directly connected nodes. In effect, all other costs are treated as infinite. What this means is that the underlying charge for transport between, say, node 1 and node 5 is made up of a charge for transport from node 1 to node 3 plus a charge for transport from node 3 to node 5.

Figure 2. Five-node network

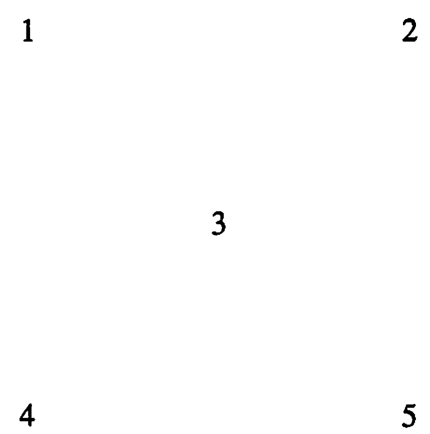

Table 3. The supply/demand pattern

Node	Supply	Demand
1	200	400
2	300	0
3	300	100
4	200	100
5	0	400

Table 4. <u>The contract cost matrix</u>

	1	2	3	4	5
1	0	p	p	-	-
2	p	0	p	-	-
3	p	p	0	p	p
4	-	-	p	0	p
5	-	-	p	p	0

The solution of the problem is straightforward and is shown in the notional flow matrix in Table 5. Note that two of the potentially available connections are empty: no flows are allocated to the connections between nodes 1 and 3 and between nodes 3 and 4. The number of non-empty cells is 4 (N-1), which together with the 5 (N) diagonal elements gives 9 (2N-1) equations for the entry and exit charges.

Table 5. <u>The notional flow matrix</u>

	1	2	3	4	5
1	200	0	0	0	0
2	200	0	100	0	0
3	0	0	0	0	300
4	0	0	0	100	100
5	0	0	0	0	0

Arbitrarily setting the exit charge at node 1 to zero yields the corresponding entry and exit charges shown in Table 6. These show that incremental transport charges are negative for some changes in the supply/demand pattern. For example, charges are negative for transport corresponding to incremental supply at node 5 and for incremental demand at node 2. Nevertheless, the total revenue yield from the tariff is positive, and equal to 7p.

Table 6. Entry and exit charges

Node	Entry	Exit
1	0	0
2	p	- p
3	0	0
4	0	0
5	- p	p

6.2 A five-zone case

The above example can be re-interpreted in terms of zones rather than nodes, although in this case it will generally be inappropriate to levy a zero underlying charge for transport from a supply point in zone i to a demand point in zone i.

One approach is to distinguish, at the first stage of the exercise, between local or intra-zonal transport charges and long-distance or inter-zonal charges. The contract cost matrix above can then be interpreted as representing the inter-zonal component of charges. As for local transport, the simplest approach is to introduce an underlying charge for transport within a particular zone, say f_{jj}, which may differ as between zones to reflect local cost factors. The total per unit charge for transport from zone i to a different zone j can then be assumed to be made up of the inter-zonal charge plus the average of the local transport charges for the two zones.[14] That is:

$$p_{ij} + f_{ii}/2 + f_{jj}/2$$

Assuming, for simplicity, that $p_{ij} = p$ for adjacent, connected zones and that $f_{ii} = f$ for all i, it is easily shown that the zonal model will give the same notional power flows as in the previous nodal model. However, the shadow prices, and hence the entry and exit

[14] This approach will be closest to the transmission cost structure when the network has a "hub and spoke" pattern. Then $f_{ii}/2$ could reflect the cost of transport to the hub of zone i, p_{ij} the cost of transport between hubs i and j, and $f_{jj}/2$ the cost of transport from the hub of zone j.

charges, will now be different because the intra-zonal transport charge leads to equations of the form:

$$u_i + v_i = f_{ii}.$$

The new set of zonal entry and exit charges are shown in Table 7. Total revenue recovery in this case is 10f + 7p, broken down as between 10f + 3p from entry charges and 4p from exit charges.

Table 7. Zonal entry and exit charges

Zone	Entry	Exit
1	f	0
2	f + p	- p
3	f	0
4	f	0
5	f - p	p

In the above example, it can be seen that all of the local-transport charges are incorporated into entry charges and that the exit charges are the same as in the nodal case. This follows from the arbitrary assumption that the exit charge for zone 1 is zero. A different assumption here will lead to a different balance between entry and exit charges, although the spatial differences in charges and the total level of revenue recovery will be unaffected.

Table 8. Alternative entry and exit charges

Zone	Entry	Exit
1	f - y	y
2	f + p - y	- p + y
3	f - y	y
4	f - y	y
5	f - y - p	p + y

To illustrate, assume that the exit charge in zone 1 is y. Then the charging structure is as shown in Table 8. Revenue recovery is now 10f - 10y + 3p from entry charges and 10y + 4p from exit charges.

For certain values of the parameters, in the zonal case it is possible that all the charges are non-negative. Thus the lowest entry charge is f - p - y and the lowest exit charge is y - c. The parameter y can be chosen to make both of these positive, provided that f - p > p (any y such that f - p > y > p will suffice). All charges can be non-negative, then, provided that f > 2p. That is, local transport charges must be sufficiently large relative to long distance charges.

7. Refinements of the basic approach

7.1 The spatial structure of charges

As already noted, the fact that, in the simple linear network example, the entry and exit charges calculated from the optimisation problem yield a true marginal cost tariff structure is due to the exact matching between transport charges and the network operator's own costs. Where the underlying charges for use of a particular part of the network deviate from the network operator's costs the equivalence will break down.

For example, suppose that there are economies of scale in transmission such that the marginal cost of transport is lower than the average cost. Then if the transport charge p is based upon average transmission costs, say, the shadow prices generated by the solution of the allocation of nominal flows problem will no longer correspond to incremental transmission costs (and nominal flows will not correspond to actual flows). Nevertheless, the spatial structure of the charges is likely to be more closely correlated with the spatial structure of the network operator's marginal costs than is a tariff structure derived from a simplistic approach to notional paths. The reason for this is that the procedure for minimising total contractual payments is based upon a network approach to the problem. Network characteristics are therefore automatically incorporated into the solution in a way that they are not in more simplistic notional flow approaches.

7.2 Non-linear charging

In conditions of economies of scale, however, there may be very strong reasons for preferring distance-related charges to be linked more closely to marginal than to average costs. One reason is the traditional efficiency argument (although, as noted in section 2, this does not imply that strict marginal cost pricing is optimal); another reason is the public policy benefit of avoiding large spatial differentials in commodity prices. Thus, where marginal costs are lower than average costs, linking underlying (inter-node or inter-zone) transport charges to the former rather than the latter will tend to lead to lesser spatial differentiation in commodity prices. Moreover, where there is some degree of choice concerning the appropriate definition of marginal costs -- which, given the subjectivity of cost estimates, there generally will be -- a policy preference for greater spatial homogeneity in commodity prices will tend to speak in favour of a choice of transport charge closer to the lower end of the possible spectrum.

The problem that then emerges is that of revenue recovery: the yield to the network operator from the use-of-system tariff may fall well short of the operator's total costs. There are, however, a number of ways in which this problem can be tackled, including non-linear pricing, of which two-part tariffs are perhaps the most familiar example.

The underlying transport charge could, for example, take the following form:

total charge for transport of q units of commodity from node i to node j

$$= F + p_{ij}q.$$

The approach then follows the lines already discussed, with, say, F/2 being added to each user's entry payment and/or exit payment.

Such a tariff structure would, however, tend to penalise small users, in the sense that each user would pay the same fixed charge independently of size. Apart from straightforward considerations of static efficiency -- an equal fixed charge for each customer is unlikely to be the least distortionary way of raising the necessary revenues[15] -- the discouragement of smaller users would likely count as a negative factor in policy making circles as a result of its potentially negative impact on competition.

The simplest solution is to link the additional charging component to the size of the transportation demand, for example via the following:

per unit charge for transport from node i to node j

---

15 Compare with Ramsey pricing rules.

$$= f + p_{ij}.$$

The additional component is now no longer fixed in relation to output, but it is independent of location. As a consequence it should have relatively little effect on location decisions and on the spatial structure of commodity prices. The procedure can be further refined whilst retaining this latter characteristic by making the first component of the unit charge a function of the size of the transportation demand. Thus it is possible to use the least-cost notional flows approach to produce a structure of charges that incorporates quantity discounts to large users, say.

It can also be noted that the above method of recovering revenue is formally equivalent to introducing uniform local transport charges, as was done in the zonal model of section 6. The arguments concerning revenue recovery that have just been rehearsed tend to suggest, therefore, that there will be advantages of a tariff structure that makes the ratio of local to long-distance charges somewhat greater than the ratio of the network operator's corresponding incremental costs. This in turn will make it more likely that a structure of entry and exit charges that avoids negative charges can be devised.

Some problems remain, of course. High values of f may create incentives for bypass of the system by customers requiring short-distance transportation. Nevertheless, as a general principle, regulatory experience suggests that, other things equal, there is, and will continue to be, a policy preference for loading the burden of revenue recovery more on to tariff components that are not distance related than on to those that are distance related.

7.3 Temporal variations in charges

The procedures described can be applied for any given time period. Thus, where transportation costs vary with time -- because, for example, of daily or seasonal variations in loads -- underlying contract charges can be made time dependent, as in familiar peak-load pricing arrangements.

To illustrate, in the simplest case the underlying charge for transport from i to j can be specified as $p_{ij}(t)$, where t indicates the time at which the transport takes place. Thus $p_{ij}(t)$ may be set lower at night times and during the summer than at daily demand peaks and during the winter. Within any given period, the notional flow allocation problem is solved using the charges relevant for that period, leading ultimately to entry and exit charges that vary over time.

7.4 Other cost factors

Some cost factors in a network can be spatially specific, but not distance related. A good example is the cost of reducing the voltage level as electrical power passes along a distribution system. A transformer will be located at a specific point of the network -- and it is therefore appropriate to take its location into account -- but the cost involved is not directly related to distance in any way.

The notional flows model can be extended to allow for this type of issue by treating the location at which voltage is altered as two separate, but connected nodes. Let these nodes be h and l, representing the higher and lower voltage sides of the transformer respectively. Then a charge p_{hl} can be levied on the flow from node h to node l, and this charge can be incorporated in an extended charges matrix to be used in allocating notional flows. The result will be an entry and exit tariff structure that distinguishes between voltage levels.[16]

8. Capacity and commodity charges

One method of dealing with the peak-load issues raised in sub-section 7.3 above is to differentiate between capacity and commodity charging components. The capacity charge is levied on peak-demand for transportation over a given period, whereas the commodity element is levied on actual demand for transportation. For example, within a particular year, a particular electricity supply contract may give rise to a power flow at time t of z(t) which has a maximum value of zm. The underlying charge would then take the form of a capacity charge kijzm and a commodity charge in each period of $p_{ij}(t)z(t)$.

Since it is total system demand for transportation that gives rise to the higher costs at the peak, the search for more cost-reflective tariff structures tends to suggest that it is not so much the maximum demand of each customer but rather the demand of each customer at the relevant system peak[17] that should form the base for the levying of capacity charges. Nevertheless, the individual maximum demand may serve as a satisfactory approximation, particularly if there are additional schemes which allow rebates for

16 Following privatisation in the UK, electricity transmission charges are differentiated by location but not by voltage, while distribution charges are differentiated by voltage but not, at least within regions, by location.

17 The problem is further complicated by the fact that the period of greatest pressure on one part of the transportation system might not coincide with the period of greatest pressure on another part.

demands that clearly do not contribute to pressure on capacity. The simplest example of the latter would be "interruptible transportation" tariffs. Thus, where the contract contains provisions for the network operator to cease transport of a given load when the system comes under pressure, it would be appropriate for that contract to carry no capacity charge.

Where the approximation based upon capacity charges levied on individual maximum demands is adopted, application of the notional flow allocation method is again straightforward. In this case there are in fact two problems to solve. First, taking "non-interruptible" maximum demands for transportation, the problem can be solved for notional peak flows using the capacity charges as the relevant components of the contract cost matrix. Second, in any given period, the problem is solved for all transportation demands, including interruptible demands, using the commodity charges as the relevant elements of the contract cost matrix. The result is a series of entry and exit charges, with capacity elements relating to maximum demand and commodity elements (possibly differentiated by time) levied on actual flows.

In practice it may be found that most of the network operator's costs are related to peak demands for transportation, and therefore that, if the underlying charges were to be based only upon incremental costs, the resulting structure would be heavily tilted towards the capacity element. Such is recognised to be the case in gas:

> "For a given peak transmission capacity requirement, increasing the annual volume would have a relatively modest impact on the costs of transmission. By contrast, increasing the peak day capacity requirement for a given annual volume would add significantly to the overall cost of transmission." (From Gas transportation: a public consultation document, issued by British Gas, March 1992).

The situation is somewhat different in electricity, where line losses play a more important role. In the UK industry, however, line losses are priced through the pooling arrangements established at the time of privatisation. The costs of line losses are not currently borne by the National Grid Company, whose own costs are very heavily driven by peak requirements.

It can be argued, therefore, that a network pricing system based only upon capacity charges might suffice. In support of this proposition, it can be pointed out that, over time, the shorter-term efficiency properties of such a system could be developed by allowing for resale of capacity. For example, an interested party requiring transportation in an off-peak period would be required to be in possession of the relevant capacity rights, but such rights could be purchased in secondary markets from another party who had excess transportation capacity at that time. Competition in the secondary market would then lead to more (short-term) cost-reflective, temporally differentiated prices for use of the network. This approach has obvious advantages in the shape of an increased role for markets and a correspondingly reduced role for regulation.

There are, however, also strong arguments for the introduction of significant commodity charges, particularly in the absence of effective secondary markets in capacity. One of the most important is that, if capacity charges are based upon incremental costs, and particularly if they are based upon an incremental cost measure that is chosen because, being low, it does not lead to very wide spatial variations in commodity prices, the revenue yield from capacity charges may fall well short of total cost recovery. Arguably, then, transportation demand throughout the whole year, rather than peak demand alone, provides the most suitable base for the recovery of additional revenues.

One point here is that total transportation demand may be a more equitable base for revenue recovery than capacity charges which, for example, bear heavily on small but very peaky demands. Another is that it may be efficient to place a significant part of the revenue recovery burden on commodity charges. To rely only on capacity charges would introduce strong incentives for network users inefficiently to alter their load profiles, particularly if interruptible tariffs were available. Such incentives can be countered by widening the base of revenue recovery to include a significant commodity element.

9. Conclusions

Although it is well known that notional path methods lead to charges for use of networks that can deviate radically from costs, the problem appears to have less to do with the concept of a notional path itself and more to do with the way in which notional paths are calculated. Typically these paths are identified by assessing the relevant load in isolation from all other loads on the transportation system.

It is, however, relatively straightforward to identify notional paths on a network basis, taking account of the interactions between transportation demands and allowing an incremental load to affect the prior notional paths of other loads. More precisely, for any given set of charges for point-to-point transport over the network, notional flows can be identified which minimise the total charges incurred by network users whilst satisfying the given demands for transportation. When a new transport demand is added it has a well-defined effect on notional flows and, hence, on total charges. The effect on total system charges can therefore be used as the basis for the individual charge levied on the system user.

The above procedure -- which has been labelled least-cost notional flow analysis --leads naturally to a tariff structure for use of the network that is based upon entry and exit charges. The procedure can be made simpler or more complex by varying factors such as the number of nodes or zones to be included and the degree of accuracy with which point-to-point charges are made to reflect the economic costs of transmission or distribution. The closer the underlying charging structure is to the actual cost conditions in the network, the closer will be the notional flows to optimised, actual flows, and the closer will be the tariff structure to that in a full marginal cost approach. However, even relatively simple characterisations of the network and highly approximate estimates of underlying transmission or distribution costs can lead to notional flows that are a great deal closer to actual flows than those derived from approaches which ignore network factors entirely. Moreover, by keeping underlying structures relatively simple, the resulting tariffs are made more transparent and easier to monitor.