Journal of Transport Geography 1994 **2**(1) 31–40

The hub network design problem

A review and synthesis

Morton E. O'Kelly

Department of Geography, The Ohio State University, Columbus, Ohio 43210, USA

Harvey J. Miller

Department of Geography, University of Utah, Salt Lake City, Utah 84112, USA

Hubs, or central trans-shipment facilities, allow the construction of a network where large numbers of *direct* connections can be replaced with fewer, indirect connections. Hub-and-spoke configurations reduce and simplify network construction costs, centralize commodity handling and sorting, and allow carriers to take advantage of scale economies through consolidation of flows. Such networks have widespread application in transportation. This paper presents a structured review of research on the hub network design problem. Three critical design questions need to be considered: (a) are the nodes in the network assigned exclusively to a single hub? (b) are direct node-to-node linkages permitted to bypass the hub facilities? and, (c) are the hub facilities fully interconnected? The nature and difficulty of the hub network design problem depends on the analyst's judgement with respect to these questions. We review analytical research papers, and give brief empirical examples of eight different network design protocols.

Keywords: hub and spoke, network design, location

Flows of people, commodities, information and energy all require a complex network of interlinkages between origins and destinations. A special kind of network, namely, the *hub* network is designed for servicing human, commodity or information flows between multiple origins and destinations, ie, the *many-to-many* distribution problem. Hubs, or central trans-shipment facilities, allow the construction of a network where direct connections between all origin and destination pairs can be replaced with fewer, indirect connections. These configurations reduce and simplify network construction costs, centralize commodity handling and sorting, and allow carriers to take advantage of scale economies through consolidation of flows (Chestler, 1985; Devany and Garges, 1972; Kanafani and Ghobrial, 1985; Toh and Higgins, 1985).

Hub-and-spoke networks are applicable to many different types of transport problem. Examples of hub-and-spoke systems include: airline passenger carriers (Shaw, 1993); overnight package delivery services (Chestler, 1985); and rail sorting yards (Bodin *et al*, 1980). While these diverse applications are very well known, the prospects for a comprehensive model for hub network optimization are limited at the moment. There are so many different approaches to the problem, it is difficult to form any generalizations. Indeed, there exists a disconcerting number of definitions and ideas about what constitutes a hub. The following paragraphs discuss briefly the various concepts of hub which have been used in the operations research literature, in air passenger transportation and in air package delivery.

In the case of the early literature on management science and operations research, the concept of hub seems to have been synonymous with a central warehouse or facility. (See Minas and Mitten [1958] where a model for scheduling truck movements in and out of a depot or hub is presented without any notion of sorting or throughput.) Thus, a hub is in essence a warehouse or a central depot, which is located at the centre of a set of demand regions. Conversely, Goldman (1969) analysed what is actually a hub facility, but referred to it as a 'center'. As noted by Campbell (1991a) Goldman located centres to minimize the sum of transport costs over a set of origin–destination pairs, and so formulated a general multiple-hub assignment problem.

In air passenger transportation, as defined by the US Federal Aviation Administration (FAA), the term 'hub' is not based on the hub-and-spoke

0966-6923/94/010031-10

switching operation which is the basis for the definitions in this paper. For the FAA the term hub is taken to mean a geographical area, classified on the basis of the percentage of total passengers enplaned in that area. For example, in the 1991 *Airport activity statistics* publication, the FAA defined a *large hub* in 1991 as an area which enplaned at least 4 283 192 passengers (ie at least 1% of total passengers). These large hubs accounted for 28 community areas, with 55 airports, and enplaned 73.16% of all passengers (see also Shaw, 1993, p. 48; and Dempsey and Goetz, 1992).

In package delivery systems, such as United Parcel Service, the hub terminology is used to denote almost all major sorting centres. The company, in 1992, had over 2250 operating facilities; of these, over 200 are identified as hubs! Clearly, however, their major *air* hubs are the kind of centre we are concerned with here. There are four such facilities: a main hub (Louisville, KY), and three regional air hubs (in Philadelphia PA, Dallas TX, and Ontario CA). In this paper, the term *hub* refers to this more specialized meaning; that is, it is used to denote a major sorting or switching centre in a many-to-many distribution system. Therefore, the key idea is that the flow between a set of origin and destination cities passes through one or more hubs, *en route* to the final destination.

The hub network design problem, as it is discussed in this paper, is a complex mixture of locational analysis and spatial interaction theory (O'Kelly, 1986). In its most general form, this problem involves: (1) finding the optimal locations for the hub facilities; (2) assigning non-hub origins and destinations to the hubs: (3) determining linkages between the hubs; and, (4) routeing flows through the network. Not only is the number of the decision variables large, but the solutions to these individual problems are highly interdependent. In practical terms, there are at least three approaches to handling the complexity. The first is to adopt a partial approach, whereby some aspects of the decision variables are simplified for mathematical convenience. An example of this strategy is the common assumption that transportation costs are independent of flow volume, despite the well-known importance of scale effects in reality (Campbell, 1990a). The second is to find a decomposition of the problem into convenient subproblems as exemplified by the division of the network into backbone and feeder subnets (see examples in Chan and Ponder, 1979; Chung *et al*, 1992). Finally, the third approach is to recognize the inherent mathematical difficulty, and to seek a local rather than a global optimum to the problem. Thus several researchers have begun to develop sophisticated mathematical programming *heuristics* for hub design (Abdinnour and Venkataramanan, 1992; Klincewicz, 1991, 1992; O'Kelly, 1987; O'Kelly *et al*, 1993; Skorin-Kapov and Skorin-Kapov, 1992).

A set of convenient but restrictive modelling assumptions can be exploited in order to manage the hub network design problem. The standard hub network topology, which we call *Protocol A*, consists of a relatively large number of nodes each directly connected to only one of a small number of completely interconnected hubs, ie, the pure 'hub and spoke' configuration. Protocol A serves as the basis for many efforts to solve the hub network design problem (eg, Campbell, 1991a, 1991b; Klincewicz, 1991, 1992; O'Kelly, 1986, 1987, 1992a, 1992b; O'Kelly and Miller, 1991; Skorin-Kapov and Skorin-Kapov, 1992). Later in this paper, we discuss variants on the hub network design problem, and we call them *Protocol B, C, . . ., H.*

Although the standard hub network topology is *convenient* from an analytical point of view, researchers have had to relax some of its restrictions in order to remain *relevant* to real-world distribution problems. In general, these extensions greatly complicate the design problem, requiring the use of additional simplifying assumptions in order to be tractable. As a result, approaches to the hub problem have become extremely non-standardized. Partly due to these disparate approaches even basic definitional issues regarding the components of a hub network are unresolved in the literature, as reflected in our discussion of varying hub definitions.

Our goal in this paper is to organize the growing literature on hub network design and provide a framework for standardizing the hub network design problem. In this paper, we review the characteristics of the hub network design problem and develop a series of design features that clearly specify the rules for constructing a particular hub network type. This framework can serve as a standard language for comparing different hub network design applications. In addition, the protocols indicate the complexity of different design problems and suggest a broad strategy for addressing these problems.

In the next section of this paper, we discuss properties of the standard hub network design problem. In the third section, we identify common departures from Protocol A restrictions in real-world hub networks and review attempts by researchers to accommodate these complexities. In the fourth section, we develop a series of hub network designs as a standard classification system for this problem. This includes a formal statement of definitional issues that have been neglected in the literature, presentation of the classification system and discussion of the system's implications for the design problem. The fifth and final section provides some concluding comments.

Hub network design under Protocol A

The standard hub network Protocol A is *defined* as the product of three simplifying restrictions: (1) all hubs are fully interconnected; (2) all nodes are connected to only one hub; and, (3) there are no direct non-hub to non-hub (internodal) connections. An example can be seen in *Figure 1*. The conceptualization of the standard hub network is similar to Aykin (1993) who refers to a network like Protocol A as a 'strict hubbing policy'.

The Protocol A design has two important properties. One property is *deterministic routeing*. Given fixed hub locations, allocations of non-hub origins and destination to hubs, and the triangle inequality with respect to distance, there is only one shortest path between any origin–destination pair in the network. Since each non-hub origin and destination is connected to only one hub and all hubs are interconnected, the triangular distance inequality means that the shortest path can be found simply by choosing the direct connections between a non-hub origin or destination and its hub and between the hubs if necessary. A second property is a *p-median problem constraint set*: Protocol A network characteristics allow the hub network design problem to be stated in similar format to a traditional optimal location problem. The location literature has in turn been a fruitful source of algorithms for the hub location problem. These two properties allow the hub network design problem to be stated as analogues to traditional location problems. *Table 1* summarizes these linkages.

Under Protocol A, the minisum (ie minimize aggregate flow cost) single-hub problem in planar space can be stated as an easily solved Weber least-cost location problem (O'Kelly, 1986). Also, the minimax (ie minimize the most costly network flow) single-hub problem in planar space can be solved as a round-trip location problem for which efficient solution algorithms exist (O'Kelly and Miller, 1991). The minisum, multihub problem in planar space can be treated as a multifacility location–allocation problem (Aykin and Brown, 1992). If distances are measured as squared Euclidean distances, convenient mathematical properties facilitate the solution of very large planar hub location models (O'Kelly, 1992b). The objective function for the minisum, multihub problem in discrete space under Protocol A can be stated as a quadratic assignment problem with constraints similar to the *p*-median problem (O'Kelly, 1987). While this latter problem is difficult to solve optimally (Aykin, 1988; Aykin and Brown, 1992; O'Kelly, 1986, 1987), several heuristic procedures have been developed. These procedures differ mostly with regard to node-hub assignment methods (see Campbell, 1991a, 1991b; Klincewicz, 1991, 1992; O'Kelly, 1987; Skorin-Kapov and Skorin-Kapov, 1992).

It may also be noted from *Table 1* that several Protocol A design problems are either trivial, or unsolved to date. In the former category are single-facility problems in discrete space: under both minisum and minimax objectives, this problem can be solved through simple enumeration. More complex, and still unsolved, is the multiple-hub, minimax problem in both planar and discrete space, although Campbell (1991a, 1991b) has introduced a number of formulations which extend covering models to the hub network design problem.

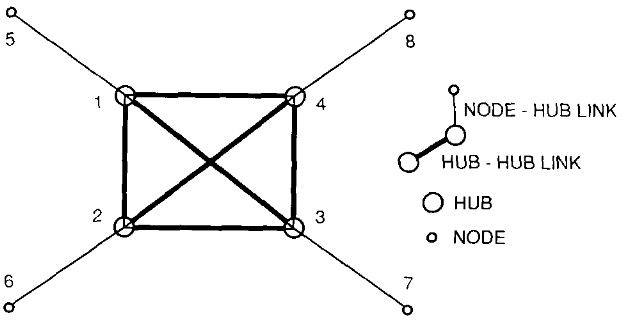

Figure 1 Example Protocol A network

Relaxing Protocol A restrictions

The generic 'hub and spoke' topology serves as the basis for the many-to-many distribution problem in a variety of empirical transport and communication applications. However, the characteristics of these real-world distribution problems have resulted in hub network configurations that typically violate one or more of the Protocol A restrictions.

Figures 2 and *3* illustrate empirical hub network applications in air and ground transportation, respectively. *Figure 2* provides the route structure (as of May 1991) for Skyway Airlines, a regional air passenger carrier based in Milwaukee, Wisconsin. Several of the network properties violate Protocol A restrictions. Internodal connections are present (eg Madison–Rockford, Saginaw–Flint, Kalamazoo–Lansing). Also evident is a feature known as a 'spider leg' (Marsten and Mueller, 1980) in which service locations are arranged in purely linear fashion (eg Peoria–Bloomington/Normal–Detroit). *Figure 3* illustrates the US route structure for Yellow Freight systems. *Figure 3a* provides the feeder (spoke) linkages to regional hubs, while *Figure 3b* indicates the interhub 'linehaul' linkages. Protocol A restrictions are violated at both network levels: several nodes are connected to more than one hub and the interhub network is not fully interconnected.

Several researchers have examined design problems for hub networks with more complex topologies than allowed hitherto. In some special cases, modification of the Protocol A restrictions actually simplifies the design problem. For example, when multiple-hub assignment is allowed, the allocation of nodes to hubs can be expressed under certain conditions as a linear assignment problem (Campbell, 1991a, 1991b). O'Kelly and Lao (1991) show that a model with allocations to both a mini- and a master-hub can be solved optimally using

Table 1 Traditional location problem analogues under Protocol A restrictions

Number of hubs	Spatial constraint	Design objective	Problem characteristics	Source
Single	Planar	Minisum	Weber least-cost location problem	O'Kelly (1986)
Single	Planar	Minimax	Round-trip location problem	O'Kelly and Miller (1991)
Single	Discrete	Minisum	Trivial; complete enumeration	
Single	Discrete	Minimax	Trivial; complete enumeration	
Multiple	Planar	Minisum	Location–allocation problem	Aykin and Brown (1992); O'Kelly (1992b)
Multiple	Planar	Minimax	Unsolved	
Multiple	Discrete	Minisum	Quadratic assignment problem with p-median constraints	Aykin (1988); Klincewicz (1991); O'Kelly (1987), O'Kelly (1992a)
Multiple	Discrete	Minimax	Integer programming formulation	Proposed in Campbell (1991b)

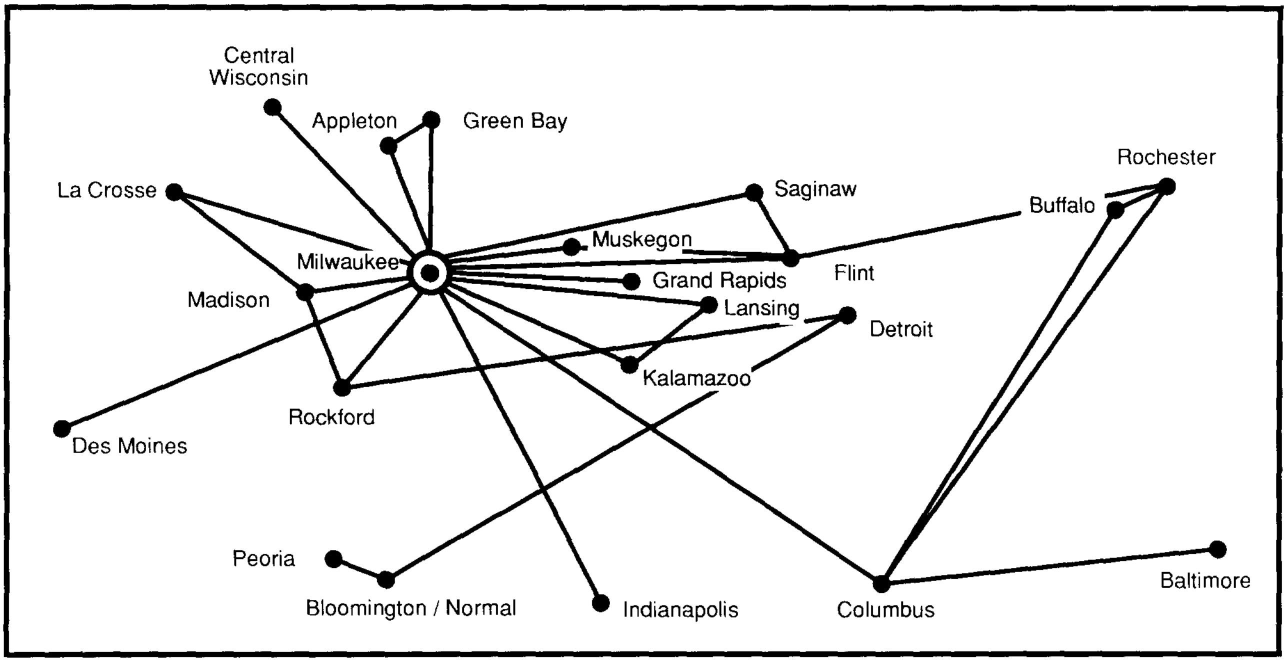

Figure 2 Skyway Airlines
Source: Columbus Dispatch (30 May 1991, p. A–2).

linear programming. In general, however, relaxing Protocol A restrictions greatly complicates the hub network design problem. This added complexity has necessitated the use of additional restrictions in order to manage the design problem. *Table 2* provides some examples.

One of the more common Protocol A relaxations attempted is the assignment of nodes to more than one hub. Multiple-hub assignment can save transportation costs by tailoring the selection of hubs to the eventual destinations of the flows being shipped from an origin node thus reducing the distance travelled. In addressing this routeing problem Daganzo (1987), and Hall (1987) are able to derive analytical solutions only by restricting the spatial dimension of the problem to linear space or L_1 metric (the latter metric limits travel to rectangular dimensions). Daganzo (1987) and Hall (1987) also fix the locations and service areas of the hubs. This restriction is relaxed by Campbell (1990a, 1990b).

Partial interhub connections concentrate flows at particular hub facilities, which allows the exploitation of flow-processing economies of scale. Leung *et al* (1990) allow partially interconnected hubs as well as multiple hub assignment by separating the node-hub assignment problem from the interhub routeing problem. The routeing problem is solved by treating it as a multicommodity flow problem, with each commodity distinguished by its origin–destination pair. Chou (1990) restricts topology to partially interconnected hubs by requiring the network to be minimally connected. Since the network is a minimal

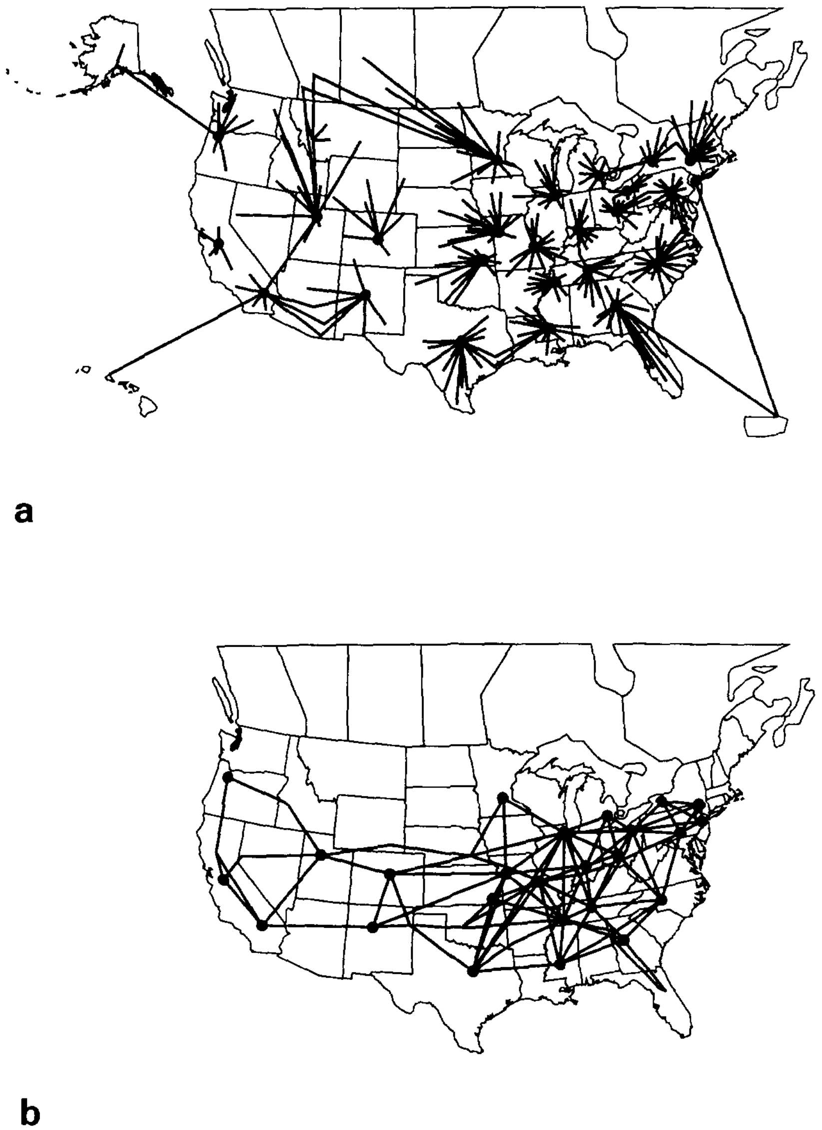

Figure 3 Yellow Freight: (a) hub and spoke, and (b) linehaul system

spanning tree, routeing can be determined through the connectivity matrix. Chou (1990) also relaxes this restriction somewhat by introducing a link-capacity constraint which can result in a more connected topology.

Internodal linkages can provide direct service between locations that have a high degree of interaction. The interesting aspect of this Protocol A relaxation is that, although the analyst may permit certain routes to be served directly, the model determines whether or not such direct routes are economically viable (see for instance Aykin, 1992). In practice, the use of direct node-to-node connections to 'bleed off' larger predictable flows from the hubs is noted in air express. It should be emphasized that in terms of our definitions, these direct node-to-node pairs do *not* create a hub at the node, as the usual hub trans-shipment functions are absent. Nevertheless, if a node is directly connected to a large number of other nodes, there would seem to be a strong case for developing full hub functionality at that location. In a slightly different vein, Flynn and Ratick (1988) allow these linkages in the form of 'stopover' air service in their air transport network model. However, the overall network is already established, and the design problem consists of feeding additional service into the established hub network. One of the most general hub network design models was formulated by Powell and Sheffi (1983). Their analysis includes both a statement of the optimal design problem and a heuristic solution procedure. The optimal version requires only one directional link to enter and leave each node or hub, but this restriction is relaxed in the heuristic solution procedure. The solution procedure is a local improvement strategy in which the user broadly directs

Table 2 Examples of hub network modelling with Protocol A violations

Source	Protocol A violation allowed	Additional restrictions	Routeing mechanism
Campbell (1990a)	Assignment of nodes to multiple hubs	Linear or L_1 space	Analytical
Campbell (1990b)	Assignment of nodes to multiple hubs	Linear space[a]	Analytical
Chou (1990)	Partial interconnection of hubs	Network required to be minimally connected subject to possible link capacity constraint	Connectivity matrix
Daganzo (1987)	Assignment of nodes to multiple hubs	Hub locations fixed L_1 space Hub service areas fixed in size and shape	Analytical
Flynn and Ratick (1988)	Internodal linkages	Limited portion of network considered	Multiobjective, hierarchical weighted covering model
Hall (1987)	Assignment of nodes to multiple hubs Hubs not connected	Hub locations fixed Limited number of hubs L_1 space	One- and two-hub routeing heuristics
Hall (1989)	Assignment of nodes to multiple hubs Hubs not connected	Hub locations fixed Limited number of hubs	One- and two-hub routeing heuristics
Leung *et al* (1990)	Assignment of nodes to multiple hubs Partial interconnection of hubs	Hub locations fixed Separates nodes assignment problem from interhub routeing problem	Multicommodity flow problem
Powell and Sheffi (1983)	Assignment of nodes to multiple hubs Hubs partially interconnected	Problem solved through user-directed local improvement heuristic with prespecified sequence of possible network changes	Not specified

Notes: [a]Euclidian space version uses Protocol A.

the search for network changes, followed by interactive modifications.

Table 2 illustrates the wide variety of restrictions exploited in order to facilitate model solutions for relatively complex configurations. Restrictions include: (i) limitations on the spatial dimension of the problem (eg Campbell, 1990a, 1990b; Daganzo, 1987; Hall, 1987); (ii) fixing selected components of the network or limiting their complexity (Chou, 1990, 1993; Daganzo, 1987; Hall, 1987; Leung *et al*, 1990); (iii) restricting the scope of the analysis (Flynn and Ratick, 1988); and (iv) partitioning the overall design problem into more manageable components (Leung *et al*, 1990; Powell and Sheffi, 1983). Thus, approaches to the hub network design problem beyond the Protocol A restrictions are disparate. This creates difficulty in comparing (and even defining) the hub network design problem across a wide range of applications.

A hub network classification system

In this section of the paper, we provide a common framework for the hub network design problem. This framework consists of formal definitions of hub network components and a classification system based on combinations of specific network design rules within the definitional parameters. We discuss the definitional issue first and then present the classification system. Finally, we discuss the implications of our system for the hub network design problem.

Addressing the hub network design problem for a wide variety of possible configurations raises some very basic definitional issues that have not been considered in the literature. However, these basic definitions are important in order to establish the basic ground rules that characterize a hub network. Without formal definitions of basic hub network components, the hub network design problem cannot be consistent across different applications.

A hub network consists of three major components: *service nodes*, *hubs* and *arcs*. A 'service node' is a point location from which flows can originate and into which only flows which are destined for that location can enter. A 'hub' has the characteristics of a service node (ie it can be a flow origin and destination) but also allows the passage of *throughflows* or trans-shipment flows which are not destined for that location. All throughflow that enters a hub must also exit that hub. Hubs are not differentiated by class or hierarchy: we assume for now that a hub can handle any amount of throughflow.

The arcs that connect the service nodes and hubs must have the following properties: (1) every service node must be connected to at least one hub; (2) a valid path must exist between all hubs. These two properties ensure that a feasible path will exist between all origins and destinations in the network. Also observe that a service node which is directly connected to all other service nodes does not influence the hub network design problem. Therefore we disregard a service node as part of the hub network if it essentially bypasses that network.

Note that property 1 does not allow the 'spider leg' configuration unless the intermediate node (eg Bloomington/Normal in *Figure 2*) is defined as a hub, since that location receives throughflow. This does not greatly reduce the generality of the hub network classification system since hubs can handle any amount of throughflow, even if this flow is from only one service node. In fact, intermediate nodes in a spider leg configuration do perform one of the major services of a hub (ie the consolidation of flows), although other services (ie sorting) may not be performed. Recently, Kuby and Gray (1993) have developed an analysis of the hub network design problem with stopovers and feeders. They suggest that in the case of Federal Express, spider leg links to the hub at Memphis are common and require careful analysis.

Since we are primarily concerned with the situation where hubs are selected from the existing set of origins and destinations (ie the discrete space problem, meaning that hubs are also flow origins and destinations), network configurations in which hubs have no interconnections (see, eg, Hall, 1987, 1989) are not valid since this would make certain hubs (as flow destinations) inaccessible from portions of the network. From the perspective of our classification system, we would consider these configurations as separate but intermeshed hub networks.

As noted earlier, the Protocol A network is the product of three assumptions: (i) all hubs are fully interconnected; (ii) all non-hub nodes are connected to only one hub; and (iii) there are no internodal (direct service node to service node) connections. Any or all of these rules can be relaxed as long as the basic assumptions discussed in this section are not violated. This provides three binary decision variables with which to define hub network types. The three decisions are:

(D1) Node assignment *either* one hub assignment,
or multihub assignment.
(D2) Direct node–node *either* not allowed,
or allowed,
(D3) Hub interconnection *either* full,
or partial.

D1 concerns whether a node can be assigned to only one hub or more than one hub and D2 refers to whether direct internodal connections which bypass the hub structure can be allowed. D3 refers to the subnetwork that connects the hubs alone; this can be fully interconnected or only partially interconnected.

The three binary decision variables create $2^3 = 8$ hub network classes, which are defined in *Table 3*. Example networks can be seen in *Figure 4*.

While it is difficult to find an exact match in the real world to these prototypical networks, there are several excellent representative examples. Protocol A is similar to the Rockwell International interplant communications system illustrated in Fotheringham and O'Kelly (1989, p. 172). Protocol B can be seen in the satellite communications network design proposed by Helme and Magnanti (1989). Note that in their approach (see p. 431, *Figure 2*), each node is connected to one of the available hubs, but these hubs are not all connected directly to one another. Instead they have a link to a super hub (satellite). Protocols C and D can be seen to a certain extent in some financial networks which have single-hub assignment, but varying degrees of direct node-to-node connections and interhub connectivity (Weinstein, 1982). Protocol E is seen in McShan and Windle (1989, p. 213) where there are connected hubs, but multiple-hub allocations. The Yellow Freight system shown in *Figure 3* is an excellent example of Protocol F. Many air passenger systems illustrated in Shaw (1993) exemplify Protocols G and H.

A key feature of our classification system concerns the complexity of the hub network design problem. The basic design questions inherent in all protocols are the locations of hubs and the assignment of non-hub nodes to hubs. Beyond this, each protocol differs with regard to the freedom to configure arcs in the network. For example, in Protocol A there are no 'free' arcs: hubs must be fully interconnected, only one arc connects any node to any hub, and direct internodal connections are not allowed. in contrast, all arcs (existing and potential) in a Protocol H network are free for variable reconfiguration within the constraints of that protocol. *Table 4* indicates the decision variables that occur in each of the design problems.

An important implication for hub network design is that while the consideration of all possible configurations for a hub network results in a very complex design problem, the use of the classification system allows this complexity to be managed to a

Table 3 Hub network classification system

	Design variables		
Design class	**Node-hub assignment**	**Internodal connections**	**Interhub connectivity**
Protocol A	Single hub only	Not allowed	Full
Protocol B	Single hub only	Not allowed	Partial
Protocol C	Single hub only	Allowed	Full
Protocol D	Single hub only	Allowed	Partial
Protocol E	Multiple hubs allowed	Not allowed	Full
Protocol F	Multiple hubs allowed	Not allowed	Partial
Protocol G	Multiple hubs allowed	Allowed	Full
Protocol H	Multiple hubs allowed	Allowed	Partial

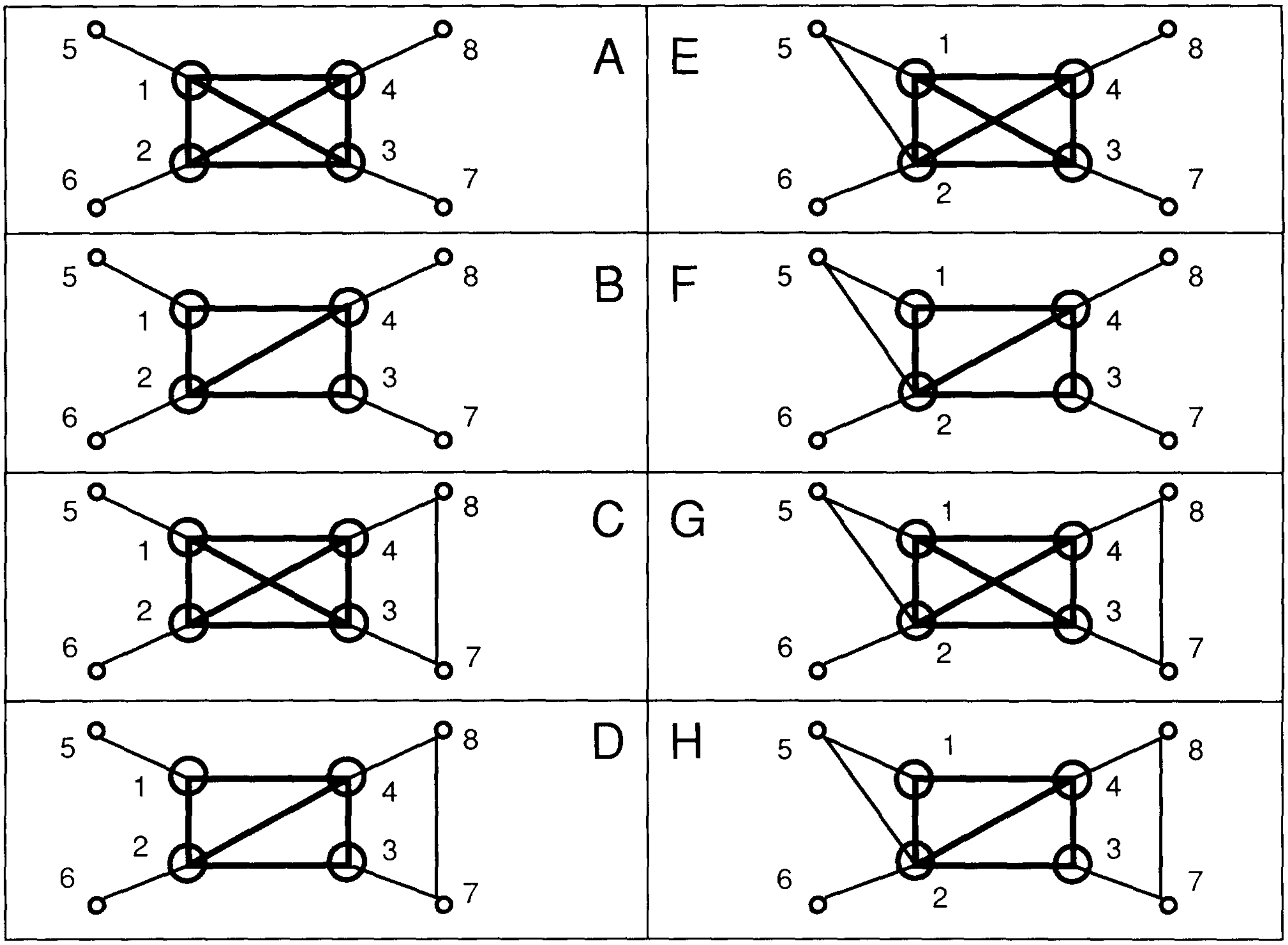

Figure 4 Example Protocol A–H networks

Table 4 Hub network design issues under different protocols with examples

Design class	Design variables	Empirical examples[a]	Analysis examples
Protocol A	Hub location Node–single hub assignment	Rockwell Interplant communications	O'Kelly (1986, 1987, 1992a, 1992b); O'Kelly and Miller (1991); Aykin (1988); Aykin and Brown (1992); Klincewicz (1991)
Protocol B	Hub location Node–single hub assignment Hub–hub links	Satellite communications	Chou (1990); Helme and Magnanti (1989)
Protocol C	Hub location Node–single hub assignment Node–node links	Financial networks	Aykin (1992, 1993); Campbell (1991a, 1991b)
Protocol D	Hub location Node–single hub assignment Hub–hub links Node–node links	Financial networks	Unknown at this time (4/93)
Protocol E	Hub location Node–multiple hub assignment	Air passenger networks	Campbell (1990a, 1990b)
Protocol F	Hub location Node–multiple hub assignment Hub–hub links	Yellow Freight	Leung *et al* (1990); Powell and Sheffi (1983)
Protocol G	Hub location Node–multiple hub assignment Node–node links	Air passenger networks	Aykin (1992, 1993); Campbell (1991a, 1991b)
Protocol H	Hub location Node–multiple hub assignment Hub–hub links Node–node links	Air passenger networks	Unknown at this time (4/93)

Note: [a]The empirical examples and references are discussed in the text.

substantial degree. The design protocols allow the decision maker to trade off the complexity of the problem against the combination of design features inherent in each network archetype. A decision maker can assess these trade-offs and choose a network type as a first approximation of the design for the particular application. Then, a specific configuration, determined according to the relatively limited number of design changes within that protocol, can be explored. For example, if a distribution problem has large amounts of interaction between some service node pairs, the decision maker may wish to abandon the convenient but restrictive Protocol A design archetype in favour of the more complex Protocol C archetype which allows internodal connections. In addition, if benefits can be derived from concentrating flows at hubs, the decision maker can also allow partial interhub connectivity by using Protocol D. However, this would make the design problem even more complex. If this complexity is undesirable, internodal connectivity could be sacrificed in favour of partial interhub connectivity by using Protocol B. Thus, the choice of a design protocol involves trade-offs between problem complexity and desired network properties relative to the particular distribution problem the decision maker is attempting to resolve.

Conclusion

Hub networks are used for solving a class of the many-to-many distribution problem. Hubs allow the construction of indirect linkages between origins and destinations, which can benefit operating costs, service provision and market position. These networks are used for air and ground transportation and communication and can have a variety of configurations. However, the design problem for all but the simplest network topologies can be extremely complex. Due to the nature of this design problem, researchers have been forced to rely on restrictive problem assumptions and have used a wide range of non-standardized approaches to the problem. In this paper, we have synthesized existing approaches to the hub network design problem and presented a framework for standardizing the problem. By doing so, we have identified examples of prototypical networks, anticipated the occurrence of other hybrid network configurations and drawn attention to some gaps in the analytical literature on these nets.

This paper has introduced the reader to the complexity of the hub-and-spoke network design problem. There are many further complexities that could be introduced. The obvious directions are to include capacity constraints (as suggested by Aykin, 1993) and to determine a dynamic facility siting plan (as suggested by Campbell, 1990b). Apart from these extensions, however, is the idea that the hub network ought to be chosen without any *a priori* restrictions on the types of connections permitted. In other words, the analyst should be able to determine the properties of the best network plan, including the level of interhub linkage, the provision of direct routes, and so on. The contribution of this paper has been to classify the types of hub network structure that *can* emerge. It remains a major research problem to offer a prescription for the best type of network for the various transport applications.

Acknowledgements

This work was supported by the National Science Foundation (SES-8821227, and DMS-9200292). A previous version of this paper was presented at the 1990 Annual Meeting of the Association of American Geographers, Toronto, Ontario. The authors thank both Dr Turgut Aykin and Dr James Campbell for providing preprints of their working papers. Thanks also to Mace Bowen, DIGIT Laboratory, University of Utah for drawing the figures.

References

Abdinnour, S. and Venkataramanan, M.A. (1992) 'Using simulated annealing to solve the p-hub location problem', manuscript, *Proceedings of the Decision Science Institute* (Abstract forthcoming)

Aykin, T. (1988) 'On the location of hub facilities', *Transportation Science*, 22, pp. 155–157

Aykin, T. (1992) 'The hub location and routing problem', manuscript

Aykin, T. (1993) 'Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem', *European Journal of Operational Research* (forthcoming)

Aykin, T. and Brown, G.F. (1992) 'Interacting new facilities and location-allocation problems', *Transportation Science*, 26(3), pp. 212–222

Bodin, L.D., Golden, B.L., Schuster, A.D. and Romig, W. (1980) 'A model for the blocking of trains', *Transportation Research B*, 14B, pp. 115–120

Campbell, J.F. (1990a) 'Freight consolidation and routing with transportation economies of scale', *Transportation Research B*, 24B, pp. 345–361

Campbell, J.F. (1990b) 'Locating transportation terminals to serve an expanding demand', *Transportation Research B*, pp. 173–192

Campbell, J.F. (1991a) 'Hub location problems and the p-hub median problem', Center for Business and Industrial Studies, WP 91–06–21, University of Missouri-St Louis

Campbell, J.F. (1991b) 'Integer programming formulations of discrete hub location problems', Center for Business and Industrial Studies, WP 91–06–22, University of Missouri-St Louis

Chan, Y. and Ponder, R.J. (1979) 'The small package air freight industry in the United States: a review of the Federal Express experience', *Transportation Research A*, 13A, pp. 221–229

Chestler, L. (1985) 'Overnight air express: spatial pattern, competition and the future in small package delivery services', *Transportation Quarterly*, 39, pp. 59–71

Chou, Y-H. (1990) 'The hierarchical-hub model for airline networks', *Transportation Planning and Technology*, 14, pp. 243–258

Chou, Y-H. (1993) 'Airline deregulation and nodal accessibility', *Journal of Transport Geography*, 1, pp. 36–46

Chung, S.-H., Myung, Y.-S., and Tcha, D.-W. (1992) 'Optimal design of a distributed network with a two-level hierarchical structure'. *European Journal of Operational Research*, 62, pp. 105–115

Daganzo, C.F. (1987) 'The break-bulk role of terminals in many-to-many logistic networks', *Operations Research*, 35, pp. 543–555

Devany, A.S. and Garges, E.H. (1972) 'A forecast of air travel and airport and airway use in 1980', *Transportation Research*, 6, pp. 1–18

Dempsey, P.S. and Goetz, A.R. (1992) *Airline deregulation and laissez-faire mythology*, Westport, CT: Quorum Books

Flynn, J. and Ratick, S. (1988) 'A multiobjective hierarchical covering model for the Essential Air Services program', *Transportation Science*, 22, pp. 139–147

Fotheringham, A.S. and O'Kelly, M.E. (1989) *Spatial interaction models: formulations and applications*, Dordrecht: Kluwer

Goldman, A.J. (1969) 'Optimal location for centers in a network', *Transportation Science*, 3, pp. 352–360

Hall, R.W. (1987) 'Comparison of strategies for routing shipments through transportation terminals', *Transportation Research A*, 21A, pp. 421–429

Hall, R.W. (1989) 'Configuration of an overnight package air network', *Transportation Research A*, 23A, pp. 139–149

Helme, M.P. and Magnanti, T.L. (1989) 'Designing satellite communication networks by zero–one quadratic programming', *Networks*, 19, pp. 427–450

Kanafani, A. and Ghobrial, A.A. (1985) 'Airline hubbing: Some implications for airport economics', *Transportation Research A*, 19A, pp. 15–27

Klincewicz, J.G. (1991) 'Heuristics for the *p*-median hub location problem', *European Journal of Operational Research*, 53, pp. 25–37

Klincewicz, J.G. (1992) 'Avoiding local optima in the *p*-hub location problem using tabu search and GRASP', *Annals of Operation Research*, 40, 283–302; presented at ISOLDE V, June 1990

Kuby, M.J. and Gray, R.G. (1993) 'The hub network design problem with stopovers and feeders: the case of Federal Express', *Transportation Research A*, 27A, pp. 1–12

Leung, J.M.Y., Magnanti, T.L. and Sighal, V. (1990) 'Routing in point-to-point delivery systems: formulations and heuristics', *Transportation Science*, 24, pp. 245–260

Marsten, R.E. and Mueller, M.R. (1980) 'A mixed-integer programming approach to air cargo fleet planning', *Management Science*, 26, pp. 1096–1107

McShan, S. and Windle, R. (1989) 'The implications of hub-and-spoke routing for airline costs and competitiveness', *Logistics and Transportation Review*, 25(3), pp. 209–230

Minas, J.S. and Mitten, L.G. (1958) 'The hub operation and scheduling problem', *Operations Research*, 6, pp. 329–345

O'Kelly, M.E. (1986) 'The location of interacting hub facilities', *Transportation Science*, 20, pp. 92–106

O'Kelly, M.E. (1987) 'A quadratic integer program for the location of interacting hub facilities', *European Journal of Operational Research*, 32, pp. 393–404

O'Kelly, M.E. (1992a) 'Hub facility location with fixed costs', *Papers in Regional Science* 71(3), pp. 293–306

O'Kelly, M.E. (1992b) 'A clustering approach to the planar hub location problem', *Annals of Operations Research*, 40, pp. 339–353

O'Kelly, M.E. and Lao, Y. (1991) 'Mode choice in a hub-and-spoke network: a zero–one linear programming approach', *Geographical Analysis*, 23, pp. 283–297

O'Kelly, M.E. and Miller, H.J. (1991) 'Solution strategies for the single facility minimax hub location problem', *Papers in Regional Science*, 70, pp. 367–380

O'Kelly, M.E., Skorin-Kapov, D. and Skorin-Kapov, J. (1993) 'An improved lower bound estimate for the hub location problem', *Management Science* (forthcoming)

Powell, W.B. and Sheffi, Y. (1983) 'The load planning problem of motor carriers: problem description and a proposed solution strategy', *Transportation Research A*, 17A, pp. 471–480

Shaw, S.-L. (1993) 'Hub structures of major US passenger airlines', *Journal of Transport Geography*, 1, pp. 47–58

Skorin-Kapov, D. and Skorin-Kapov, J. (1992) 'On tabu search for the location of interacting hub facilities', *European Journal of Operational Research* (forthcoming), manuscript, W. A. Harriman School for Management and Policy, SUNY Stony Brook

Toh, R.S. and Higgins, R.G. (1985) 'The impact of hub and spoke network centralization and route monopoly on domestic airline profitability', *Transportation Journal*, 24(4), pp. 16–27

Weinstein, S.B. (1982) 'A perspective on financial industry networking', *Journal of Telecommunication Networks*, 1(4), pp. 317–332