

CHAPTER 3: HUBBING THEORY

3.1 Introduction

Button et al. (2002) state that in order to minimise costs and keep airfares down, airlines need to keep aircraft in the air for the longest possible time to achieve the highest possible load factor, and to coordinate their aircraft, crew and maintenance schedules. To achieve this, many airlines operate hub-and-spoke (H&S) networks which entail consolidating traffic from a diverse range of origins, destined for a diverse range of final destinations at hub airports. In the airline industry, one of the most striking changes precipitated by deregulation in the US has been the restructuring of carrier networks from a mostly linear to an H&S structure, because of the major cost reduction incurred in these networks (Levine, 1987).

This chapter deals with hubbing as a cost-minimising option for airlines and route networks. Literature relevant to the effects of hubbing and cost-effective methods of carrying out hub network design is investigated, and finally, the methodology for designing an H&S network applicable to the Africa air network is developed.

3.2 Hub Classification

Hubs are defined as collection points that serve the purpose of consolidating traffic flow. The concentration or consolidation of flow can reduce movement costs (i.e. transportation or transmission) through economies of scale, even though the distance travelled may increase (Campbell, 1996). Hubs are usually found with air networks, mail delivery systems and in telecommunications.

Hubs can be defined in two general ways: one denoting whether an airport represents a hub within a carrier-independent system of air transport (i.e. airport level) and the other denoting its role within a carrier-specific network (i.e. airline level). In the analysis of hubbing, the definition of what constitutes a hub becomes crucial (Schnell and Huschelrath, 2004). For example, O R Tambo International Airport in South Africa is a hub at airport level for a number of movements within Africa and in between continents, i.e. Australia and America, while it also acts as a hub at airline level for South African Airways, the national flag carrier for South Africa. Empirical studies differ in the criteria used to define what constitutes a hub (Button et al., 2002). Table 2 shows that hubs have various definitions, depending on the function they perform, and also shows some of the classification of hubs in their specific categories.

For the purposes of this study, a hub will be defined by its route structure, i.e. its function as a distribution point for air travel to and from its surrounding catchment area, with connecting services, irrespective of the number of originating passengers.

Table 2: Various functional definitions of airline hubs

Functional definition	Explanation	Examples
<i>Scope</i>		
Operational	Merely an operational base No connecting services offered	London-Stansted (Ryan Air, Easy Jet) Frankfurt-Hahn (Ryan Air)
Marketing	Operational base Connecting services	London-Heathrow (British Airways) Frankfurt, Munich (Lufthansa)
<i>Route structure</i>		
Hinterland	Hub serves as a distribution point for air travel to and from its surrounding catchment area	Chicago (American Airlines) Dallas (American Airlines)
Hourglass	Interface between short- and long-haul flights Directionalised routing (e.g. north-south, east-west)	Vienna (Austrian Airlines) Helsinki (Finn Air) Madrid (Iberia)
<i>Strength of local market</i>		
Weak	Relatively few originating passengers	Amsterdam (KLM) Reykjavik (Iceland Air)
Strong	Relatively many originating passengers	London-Heathrow (British Airways)
Category	Explanation	Examples
<i>Size</i>		
Primary	Most important airport of an airline Focus on intercontinental traffic (if applicable)	London-Heathrow (British Airways) Frankfurt (Lufthansa) Munich (Lufthansa)
Secondary	Second most important airport of an airline Focus on intercontinental traffic (if applicable)	

Source: Schnell and Huschelrath (2004)

3.3 Advantages of Hubbing

3.3.1 Economies of traffic density

Economies of scale (which in transport refer to traffic density) occur when the average unit cost of production declines as the amount of traffic increases between any given set of points served (Barla and Constantos, 2000). The usual argument is that an H&S network, through increased traffic density on the links to the hub (the spokes), allows airlines to use larger, more efficient aircraft and to spread the fixed costs over more passengers, thus exploiting economies of scale. Besides empirical evidence of improved returns from traffic density, other empirical studies underscore the cost advantage of hubbing. McShan and Windle (1989) suggest that a 10% increase in hubbing is associated with a 1.1% decline in unit cost, all other costs remaining equal. The technical distinction between economies of scale and scope can be seen by reference to Equation 1 where C denotes cost and Q is output; economies of scope are assessed as follows (Button et al., 2002):

$$S = \{ [C(Q^1) + C(Q^2)] - C(Q^1 + Q^2) \} / \{ C(Q^1 + Q^2) \} \quad \text{Equation 1}$$

Where:

$C(Q^1)$ = the cost of producing Q^1 units of output one alone

$C(Q^2)$ = the cost of producing Q^2 units of output two alone

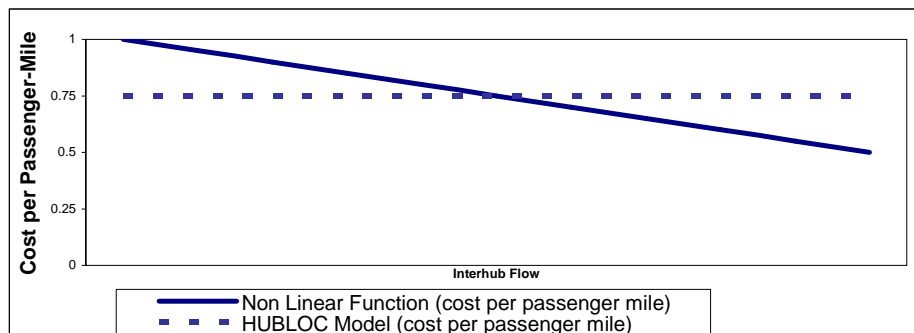
$C(Q^1 + Q^2)$ = the cost of producing Q^1 plus Q^2

Economies of scope exist if $S > 0$, while economies of scale exist if S falls as Q expands.

Furthermore, airlines introduce cost savings on indirect routes, which lead to profit-maximising prices that are well below those of direct routings, such that the lower price may over-compensate for the disutility of longer travel times and the inconvenience of changing planes (Wojahn, 2001).

The generalised cost, which is defined as the overall cost of making a trip, including all ‘time costs’, will also involve a number of non-transportation costs that are influenced by the quality of transportation provided. For example, more frequent air services from an airport reduce the likelihood that a traveller will have to bear the financial and time costs of an overnight stay on routes with low service frequencies. High frequency of services also means there is less ‘down time’ wasted as participants in international business meetings wait for fellow attendees from less busy routes to arrive (Button et al., 2002).

O’Kelly et al. (1996), in analysing the effect of an increase in discounts on hub links as passenger flow varies, found that for an H&S network, costs increase at a decreasing rate as passenger flow increases. As shown in Figure 10, with the non-linear function effect in their model, agglomeration of flow provides a benefit in that the rate at which the per-mile travel costs increase, decreases as flow increases, unlike the conventional hub location model (HUBLOC) which implies that as the flow on hub links increases, the discount stays constant.



Source: O’Kelly et al., 1996

Figure 10: Costs per passenger-mile non-linear cost function and HUBLOC model

From the above literature studies, it appears that through the lower costs of travel realised by increased flow on hub links, the benefits of traffic density in an H&S network are achieved through economies of scale (S) derived when output (Q) increases.

3.3.2 Quality of service

Hubbing offers higher flight frequencies and thus better-quality service and consumer value, necessary qualities used to measure prolonged customer satisfaction and on-going propensity to utilise products and services (Schnell and Huschelrath, 2004). The existence of the economies of scale (market) provides the consumer with a larger set of services to choose from (non-stop or with a connection), generated by greater traffic flow for the carriers (Button et al., 2002).

Button et al. (2002) found that even though the 1978 US Airline Deregulation Act contained provisions for financial support services to smaller communities (the Essential Air Services Program), the hub-and-spoke operations that came about as a result of deregulation actually stimulated the provision of services to smaller communities. Funnelling traffic through hubs makes it viable to offer higher-quality services to many smaller communities.

3.3.3 High average yield

Hubbing allows airlines to have a high average yield due to a wider 'market power', which is the ability of a market participant to control sufficient/essential facilities, to set prices profitably above, or reduce supply below, those which would occur in a fully competitive market (Schnell and Huschelrath, 2004). It was also found that hubbing was a preferred option once an airline's size and network structure had grown to a certain scale. The fact that aircraft are full or close to full with a blend of passengers with various elasticities of demand means that the airline can engage in very sophisticated demand-management and pricing schemes, effectively micro-managing the yield from the contents of the flight, based on the passengers' ability and willingness to pay (O'Kelly et al., 1996).

3.3.4 Better capacity allocation

Barla and Constantos (2000) show that hubbing has the added advantage of better allocation of capacity under demand uncertainty. Hubbing by pooling passengers from several markets into the same plane allows the firm to adjust the allocation of capacity once the demand conditions are revealed. This flexibility means that if the demand in one market turns out to be low, thereby creating excess capacity, the firm can increase sales in other markets. Moreover, if the demand in one market ends up being high with consequent binding capacity constraints, especially during peak seasons, hubbing allows a more profitable allocation of capacity since the firm can first price out the low-value travellers on several markets before eliminating travellers with higher willingness to pay.

3.3.5 Marketing advantages

Hub networks for airlines need little effort in marketing because airlines are readily associated with flights to and from the countries whose names they carry, such as British Airways, Air France, Alitalia, Austrian Airlines and Japan Airlines. It needs little marketing effort and few out-of-pocket expenses to inform a potential customer residing in California about the direct airline services of British Airways for a flight to London, Heathrow.

Besides increased production efficiency, Nero (1999) adds that an airline with a large presence in a hub airport gains significant customer loyalty advantages through marketing devices such as frequent flyer programmes and travel agency commission overrides. The existence of such marketing devices, combined with the fact that travellers value H&S network characteristics (higher frequencies of service, more connection points and a wider variety/selection of destinations), allows an H&S airline to exercise some monopoly power at the hub airport.

3.3.6 Stimulation of job creation

Hubs stimulate job creation, especially in the US high-technology sector. Statistical calculations done for 56 hub airports in the US indicated that having a hub airport in a region improves the economy through the employment

of more than 12 000 personnel. This does not mean that all hub regions benefit by this amount, but it was an average calculated across hub cities. What these econometric calculations show is that any increase in air passenger traffic at hub airports has a positive effect on employment in the surrounding metropolitan areas (Button et al., 1999). This is because traffic at hub airports in general is higher than traffic at other airports; it can therefore be inferred that hub airport cities accrue greater economic benefits than non-hub airport cities. Peeters et al. (2001) also indicate that H&S networks have a higher impact on the global environment than point-to-point networks, and that a hub airport has a higher impact on the local environment in terms of infrastructure development.

3.4 Disadvantages of Hubbing

The disadvantages of hubbing that have been identified in the literature are given below.

3.4.1 Additional running costs

A direct and non-stop flight is difficult to beat economically. The additional cost of landing and handling at an intermediate point is avoided and, more importantly, it does not add to unproductive ground time of the aircraft and crew. On product appeal, the seasoned traveller prefers the most direct itinerary, non-stop if possible, without a change of aircraft and flight at an intermediate station. H&S operations increase route frequencies, which in turn negatively affects airline costs (extra fuel consumption, extra cruise time, extra fixed costs associated with take-off/landing operations, etc.). Most fares, especially the long-haul fares, reflect the reduced route cost of the wide-body jets and their extended operating range, often allowing non-stop operation. It is therefore reasonable to assume that there is little or no margin left for a fare to cover the extra cost of the additional travel sector, especially when the through fare undersells the local fare (Zollinger, 1995).

Zollinger (1995) also points out that in order to secure a market share of traffic beyond one's national gateway, which is usually the case with H&S networks outside their countries' boundaries, like in the European Union, one needs to use costly and disproportionate efforts in advertising, promotion and solicitation activities, such as canvassing and servicing the necessary distribution channels, leading to varying operating costs and thus increasing the marginal cost price.

3.4.2 Additional travel time

Button et al. (2002) show that the need to go via a hub imposes additional costs on a traveller in terms of actual travel because of the added segment lengths involved and the transit time spent at the hub. In a hub network, direct flights do not exist, except if a passenger's final destination is the hub at which the aircraft first lands. This means that hubbing inconveniences passengers by adding extra travel time through the hubs and the transit time at hub airports before passengers reach their final destination. As shown in Figure 11, a passenger originating from A has to go through three extra sectors – A-B, B-C and C-D – before reaching destination D, whereas a direct flight – A-D – would shorten the journey.



Figure 11: Travel time added by hubbing

3.4.3 Unfair monopoly on routes

Hubbing tends to discourage entrants into a hub market, especially for a route where the rival has a hub at one endpoint and hub-to-hub routes. In the US especially, new entrants usually leave the market after a fare war when the route in question offers service to another carriers' hub (Schnell and Huschelrath , 2004). Button et al. (2002) disagree with this point, stating that some smaller carriers that entered the US airline market managed to find a niche for themselves by offering a particular kind of service, such as low-cost carriers, or by avoiding direct competition with a major carrier or, conversely, by tying in with a major carrier (notably a regional carrier).

Due to the fact that an airline operating within a hub network has a frequency advantage on a route, it enjoys a fare advantage. For example, Air France, Lufthansa and Swiss and most airlines in the US are found to charge a hub premium, making average fares higher (if at least one end-point is a hub) by an average of 4%.

3.4.4 Congestion at hub airports

For airlines there is a restriction to expansion at congested hub airports due to lack of slots in which planes can land. As a result, there is reduced flexibility on scheduling, which increases susceptibility to delays in emergency situations (Schnell and Huschelrath , 2004). Conversely, Button et al. (2002) argue that larger hub-based carriers enjoy economies of market presence and can offer more efficient network services because of scope, scale and density advantages and therefore have a greater incentive to press for additional infrastructure for runways, gates and slots at the hub airports.

Airlines that provide connecting services that flow through hub airports schedule their flights to arrive and depart in 'banks', which are periods of time in which many planes arrive and depart over a short time-span to facilitate connections. This inevitably means that there are considerable numbers of both passengers and aircraft congregated at the hub during each of these banks (Button et al., 2002). Such congestion does pose problems for fliers, who find themselves at a crowded facility, and for the airlines, which have to get their planes turned around to meet schedules. At hub airports where one carrier has a very significant amount of the traffic, the congestion costs are borne largely by its own operations and by its own passengers. In economic terms, the airline internalises the congestion costs of its interactive activities and passes them on to the passengers as levies within their fares – to the detriment of the passengers.

3.4.5 Limiting of competition

Hub carriers limit competition through the excessive market power enjoyed at their hubs because they are free from competitive pressures. Button et al. (2002) argue that the interest, which is the real test of competitiveness, is the degree of choice available to customers between their origin and desired destination. Increased levels of competition arose after the 1978 deregulation in the US market when there were an unsustainable number of new market entries at the route level as new airlines and incumbents experimented with services in the new open-market environment. Furthermore, hub carriers do face competition from specialised airlines, such as low-cost carriers (LCCs), and from technology changes, such as the preferred use of regional jets and new aircraft like the 'extended range' aircraft that fly over longer distances without the need to refuel. All these factors encourage competition on routes in hub networks.

Zollinger (1995) concludes that hubbing cannot be relied on to provide a lasting solution. It falls short of the main objectives for an airline's long-term success, which include economy of operation and product appeal. Airline planners would be well advised to look for ways to adjust the capacity offered to the genuine demand for scheduled air travel. A solution is at hand, however, since now smaller aircraft are on offer for operating short- or long-haul flights without sacrificing comfort for economics.

3.4.6 Environmental costs implication of hub networks

Research work has been done to calculate the noise and emissions in the air transport industry. The effect of H&S networks on the environment has been an area of growing concern. This is because H&S networks are characterised by longer travel distances through hubs and higher frequencies.

The social cost impact of the noise and emissions from the routes and networks in which hubs were bypassed was found to be significantly lower than that of the H&S networks. The differences in the environmental costs per passenger (noise and emission costs) ranged from 25% to 71%. This was found to be dependent on the concentration of population around the airports and the degree to which the hub routing involves extra mileage (Morrell and Lu, 2007)

3.5 Airline Hub Network

Campbell (1996) defines a hub network as one that includes nodes to represent the origin, destination and hub locations, and arcs to carry the flow. Such a network provides connections between the origins and destinations by routing flow via hub facilities. This reduces the number of arcs required to connect all origins and destinations, and it concentrates flow on these links. The creation of hubs in an air network entails designating specific airports as hubs and all the other airports as nodes in the network. The most important factors in transportation hub networks include the flow cost of transportation, i.e. moving the freight or people between origins and destinations, and the paths that the passengers will have to travel on these routes.

3.5.1 Effectiveness of hub networks

Hubs in the US succeed mostly between secondary points where no direct flights are easily available because a customer can then accept a routing through a hub and the airline can attempt to charge the full cost of each sector flight (Zollinger, 1995).

Evidence was found in the US implying that the network concentration leads to lower costs only if the carrier in question operates a large network. A survey analysis was then carried out in which a questionnaire was developed to determine airline managers' assessment of the effectiveness of H&S networks in other areas (Schnell and Huschelrath, 2004). This survey covered airlines having their home base in one of the four liberalised markets, that is Australia/New Zealand, Canada, the EU¹ or the USA. The results of the analysis and possible reasons given for the effectiveness or otherwise of H&S networks are summarised in Table 3.

Table 3: Number of hub airports categorised by size of hubs, region²

Region	Greatest number of hub airports in 2002	Destinations served within the region	Gini coefficient measuring average connectivity	Possible reason
USA	19 7 5 4 16	20-29 30-39 40-49 60-69 ≥70	0.5854	Liberalisation occurred earlier, so airlines and hub airports have been well established since deregulation in 1978.
EU	19 13 6 4	20-29 30-39 40-49 50-59	0.5557	Hubbing imposes additional travel time on passengers and this competes with the efficient road and rail network within the region.
Canada	2	20-29	0.4000	The geographical location of hub airports and competing alternative modes of transport are disadvantages.
AUS/NZ	2 1	20-29 30-39	0.4360	The geographical location of the cities served offers few benefits.

Source: Schnell and Huschelrath (2004)

The results given in Table 3 show that the difference in the number of hub airports is due to the difference in size of the regional market. However, the Gini coefficient, which is defined as a measure of statistical dispersion, is used to measure the effectiveness of hub airports in serving the various destinations for each of the regions. The hub networks of US airlines and EU airlines were calculated to have the highest coefficients since their airlines serve more routes from relatively few hub airports. This implies that there are more routes with a hub at one of their end points in the EU or the USA than in the other two regions, namely Canada and AUS/NZ (Schnell and Huschelrath, 2004). This suggests that there is a legitimate question regarding the potential efficacy of H&S systems in Africa – a question that this study will try to address.

3.5.2 Hub location

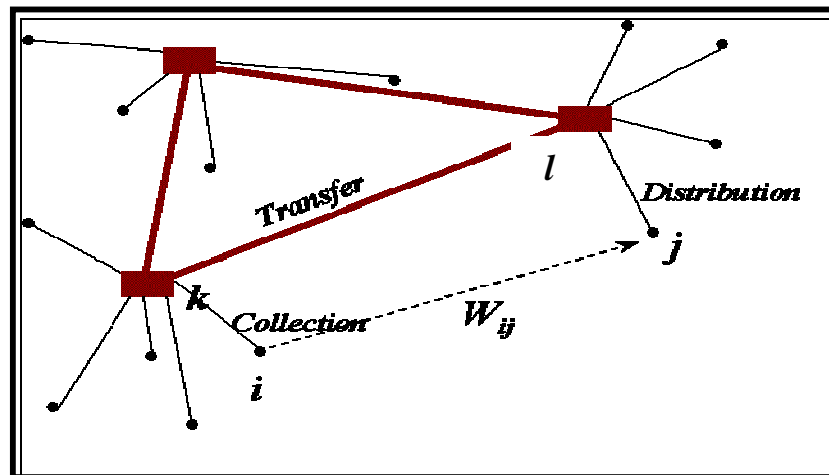
Boland et al (2004) define the hub-location problem as one concerned with creating hub-and-spoke networks, which involves locating hubs and assigning non-hub nodes to hubs with the objective of minimising transportation costs across the network. The basic information available in hub-location problems is a set of n nodes that need to exchange a known amount of flow, W_{ij} (passengers), between each pair of nodes, i and j .

¹ Including Norway and Iceland, which are both a part of the European economic area to which liberalisation of European air transport applies.

² Schnell's own calculations are based on OAG data, where the number of destinations refers to airports served by the hub operator itself or on behalf of the hub operator. This calculation is based on an operational view of hubs (Schnell, 2004).

While the simplest method of achieving this would be to connect each pair of nodes directly, this is too inefficient in a hub network and therefore all communication occurs by routing the flow via a set of hubs. The location of the hubs must be chosen from among the original set of nodes which act as collection, consolidation, transfer and distribution points, such that transferring flow between hubs is cheaper than the cost of moving flow to and from non-hub nodes.

It is usually assumed that the hubs are fully interconnected and any non-hub node can be connected directly to a hub. Note that in some cases this may require pre-processing by calculating the shortest paths through an underlying transportation network. With these assumptions and restrictions, all flow must then be routed via one, or at most two, hubs. In general, we write that flow from i to j goes via hubs k and l , where k and l could be identical if the flow is via only one hub and similarly $i = k$ if i is itself a hub or $i = j$ if j is a hub. The transportation cost as shown in Figure 12 for this connection consists of a collection cost from i to k , transfers costs from k to l and distribution costs from l to j . Some of these costs may be zero if two of these nodes are identical. Furthermore, in most applications the cost would depend on the geographical distance between the nodes in some way.



Source: CMISRO (2003)

Figure 12: Hub network flow

3.5.3 Capacities

The CMISRO (2003) defines airport capacity as the limit on the amount of flow being collected by non-hub nodes from hub nodes. Capacity is important at hub airports because of the congestion that can arise at such airports due to the limitations in facilities (in terms of gates, runways and hangars) that are realised when an airport becomes a hub. Airport capacity is of concern when there is need to consider an alternative route or a direct flight between i and j if this will cause the capacity of the hub airport to be exceeded when flow is consolidated.

3.5.4 Flow thresholds

The CMISRO (2003) defines flow thresholds as the minimum flow that is needed on some or all of the links. The flow thresholds for each of these hubs could be taken into consideration, so that the flow carried would correspond to the smallest plane operated by the company.

3.6 ρ -Hub Median Problem

The ρ -hub median problem is explained as the situation, when designing a network, where a fixed number of nodes (ρ) are allocated to be hubs and the remaining nodes have to be allocated to one or more of the chosen hubs in such a way that the operating costs of the resulting network are minimised. There are many methodologies through which the ρ -hub median problem can be solved and the study will look at the most common methodologies that have been used.

3.6.1 Single Allocation ρ -Hub Median Problem (USApHMP)

The most widely studied variant is known as the Uncapacitated Single Allocation ρ -hub Median Problem (USApHMP). In this variant exactly ρ hubs must be allocated among the n nodes and each node is allocated to only one hub (CMISRO, 2003). Analytical research on the hub-location problem began when O’Kelly (1987) devised a mathematical formulation of the problem defined as follows: Given n interacting nodes in a network, the flow between pairs of nodes i and j is denoted by W_{ij} ($W_{ii} = 0$ by assumption), while the transportation cost is denoted by C_{ij} (unit of flow between nodes i and j , with $C_{ii} = 0$) and ρ is the number of hub facilities to be located ($\rho < n$). The hub-hub discount α , on the costs of flow C_{km} for the hub-hub link is assumed to apply to all hub-hub links in the network regardless of the differences in the flow travelling across them. The problem involves finding the location of the hub facilities and thereafter assigning the nodes that minimise the total transportation cost. The first integer program formulation was proposed by O’Kelly (1987) for USApHMP, using a quadratic objective function. This problem is formulated as given below:

$$\text{Minimise} \quad Z = \sum_i \sum_j \sum_k \sum_m W_{ij} [X_{ik} X_{jm} (C_{ik} + \alpha C_{km} + C_{jm})] \quad \text{Equation 2}$$

Subject to

$$\sum X_{ik} \leq (n - \rho + 1) X_{kk} \quad \text{for all } k, \quad \text{Equation 3}$$

$$\sum X_{ik} = 1 \quad \text{for all } i, \quad \text{Equation 4}$$

$$\sum X_{kk} = \rho, \quad \text{Equation 5}$$

$$0 \leq X_{ik} \leq 1 \text{ and integer for all } i \text{ and } k$$

Where:

n	=	the number of nodes in a network
ρ	=	the number of hubs to be located
α	=	the hub-hub discount $0 < \alpha < 1$
W_{ij}	=	the amount of flow travelling between i and j
C_{ik}	=	the per-unit cost of travelling between i and k
X_{ik}	=	1 if node i is allocated to hub k , 0 otherwise

The object function in Equation 2 minimises the total network cost. The constraint in Equation 3 requires a hub to be open before a node is assigned to it. Equation 4 constrains each node to be assigned to a single hub. The constraint in Equation 5 requires that ρ hubs be open. The quadratic solution is the easiest to understand even though it is not useful for obtaining solutions directly, especially for larger networks which are more complex.

3.6.1.1 Heuristic algorithms

Due to the quadratic nature of the hub-location problem, heuristics were then used in the hub-location and allocation methodology for larger, more complex networks, with more than 25 nodes but fewer than 95 nodes in a network, to derive a single solution (Bryan and O’Kelly, 1999). Heuristics describe a set of rules developed to attempt to solve problems when a specific algorithm cannot be designed.

A variety of heuristic algorithms have been derived and researched for various hub-location problems and this research has been outlined by Bryan and O’Kelly (1999) as follows:

- O’Kelly (1987) developed the first two heuristics that computed upper bounds with the optimal objective function value for the single-assignment model.
- Aykin (1990) used flow-based assignment rather than the nearest-hub approach used by O’Kelly (1987).
- Research was then carried out to tighten the upper bounds and bring us closer to the true optimal solution: Campbell (1996) used specialised heuristics; Abdinnour-Helm et al. (1992, 1993), Aykin (1995), Ernst and Krishnamoorthy (1996, 1998) and Smith et al. (1996) used heuristics borrowed from physical sciences such as simulated annealing.
- Similarly, lower bounds for tightening the USA ρ HMP were researched by O’Kelly (1992) and O’Kelly et al. (1995), and a numerical comparison of many of these all these heuristics was done by O’Kelly et al. (1996).

3.6.1.2 Tabu Search

Tabu Search (TS) is an iterative search procedure that moves from one feasible solution to another; it is used mainly to allocate appropriate hubs to a network. Klincewicz (1991, 1992) used clustering/greedy exchanges and TS to allocate nodes to hubs. In this procedure, after a move has been made it is classified as forbidden (“Tabu”) for a certain number of iterations in the future. The primary purpose of assigning a Tabu status is to prevent cycling and to pick the optimal solutions by localising the search.

3.6.1.3 Genetic Algorithms

Genetic Algorithms (GAS) is a search algorithm used for finding the near-optimal solutions in large spaces. It was inspired by population genetics, using the mechanics of natural selection and natural genetics. The GAS method has been adopted for many operational research problems such as scheduling problems and the “travelling salesman” problem, and has been applied to location-allocation problems like the USA ρ HMP by Topcuoglu et al. (2005).

3.6.1.4 Hybrid heuristics

This form of heuristic was used by Abdinnour-Helm (1998), who applied a hybrid of Genetic Algorithms (GAS) and Tabu Search (TS) to create a model formulation called GATS. In this method a combination of the strength of GAS is used to solve the first level of the UHP-S (selecting the number and the location of the hubs), by diversifying the search, and the strength of TS is applied to solve the second level (assigning the spokes to the hubs), by narrowing down the search in a model for USA ρ HMP.

3.6.1.5 Linear programming

Bryan and O’Kelly (1999) use linear programming in hub-location research, employing the linearised version of the quadratic Equation 2 to locate and allocate hubs. Campbell (1994b) allows the use of linear programming to provide integer solutions even though it is restricted to small networks. A little later Skorin-Kapov et al. (1996) achieved a tight linearised version of the same hub-location problem, without forcing integrality, through the use of integer programming, such that exact solution values of costs were obtained for small-sized problems of up to 25 nodes.

3.6.2 Multiple Allocation ρ -Hub Median Problem (UMApHMP)

In the Multiple-Allocation Hub-Location Model (UMApHMP) originally formulated by Campbell (1994b), each interacting pair is allowed to utilise the hub that will result in the lowest travel costs for a particular origin to destination path. This implies that any single non-hub node may be allowed to interact with more than one hub if in doing so it results in lower total network costs. The UMAPHMP problem is well explained by the tight linearised version of the UMAPHMP model shown below, which was derived by Skorin-Kapov et al. (1996) and is referred to as HUBLOC. The objective function derived minimises the total network cost.

$$\text{MIN } \sum_i \sum_j \sum_k \sum_m W_{ij} (C_{ik} + \alpha C_{km} + C_{mj}) X_{ijkm} \quad \text{Equation 6}$$

Subject to

$$\sum_k Z_k = \rho \quad \text{Equation 7}$$

$$\sum_k \sum_m X_{ijkm} = 1 \quad \text{for all } i \text{ and } j \quad \text{Equation 8}$$

$$\sum_m X_{ijkm} - Z_k \leq 0 \quad \text{for all } i, j \text{ and } k \quad \text{Equation 9}$$

$$\sum_k X_{ijkm} - Z_m \leq 0 \quad \text{for all } i, j \text{ and } m \quad \text{Equation 10}$$

Where:

- a = hub-hub discount
- W_{ij} = the amount of flow between i and j
- Z_k = 1 if node k is a hub, 0 otherwise
- X_{ijkm} = the proportion of flow from i to j that is routed via hubs k and m , respectively
- C_{ik} = travel cost between i and k

This model simultaneously determines which nodes will serve as hubs and allocates non-hub nodes to the hubs. The objective function shown in Equation 6 minimises total network cost (as in the quadratic model). The constraint in Equation 7 requires that p hubs be open. The constraint in Equation 8 requires that all flow be routed via exactly one path; this means that every interacting pair (i, j) is allocated to a path via hubs k and m . The constraints shown by Equations 9 and 10 prohibit flow from being routed via a node that is not a hub. All flow must travel through at least one hub such that no direct connections are allowed between two non-hub nodes.

The model is computationally difficult to solve and, to date, optimal solutions are known only for very small networks (up to 25 nodes). The single-assignment model may be seen as a special case of the more general multiple-assignment model, since the optimal solution to a multiple-assignment model may result in single allocations for all nodes. For example, when the cost of travel across the hub-hub links is free, both the single and multiple-assignment models generate the same network design (O’Kelly et al., 1996).

Bryan and O’Kelly (1999) outline the variations of the UMA ρ HMP linearisation proposed by Ebery et al. (1998), who showed how the multiple-assignment problem may be modelled as a multiple commodity flow problem, while Klincewicz (1996) developed a heuristic for multiple assignment based on dual ascent and dual adjustment techniques for uncapacitated facility-location problems.

3.6.3 Shortest paths

In this method, the allocation problem of collecting and distributing flow can be solved by finding the shortest path between each pair of nodes in the directed graph, allowing collection from any node to any hub, transfer between hubs and distribution from any hub to any node. Ssamula (2006) proved that in route networks, the shortest path usually implies that the costs on the route are minimised, because of the ability to fly smaller aircraft which are cheap to operate on these routes.

Ebery et al. (1998) and Ernst and Krishnamoorthy (1996, 1998) solved the allocation problem involved in large numbers of possible sets of hub locations by using the all-pairs, shortest-paths method employed for multiple-allocation problems. The general methodology developed for finding the shortest paths is outlined below.

1. Partition the set of nodes into a number of clusters. In order for the lower bound to work well, these nodes should be geographically close together (i.e. with relatively small distances between them).
2. Assume we do not know the exact location of the hubs but only the number of hubs within each cluster, without knowing where in the cluster the hubs are located.
3. Calculate the shortest paths in a directed graph containing:
 - a. Collection arcs from all nodes to any node in a cluster containing at least one hub.
 - b. Distribution arcs from any node in a cluster containing at least one hub to all other nodes.
 - c. Transfer arcs between nodes in different clusters if they each have at least one hub.
 - d. Transfer arcs between nodes in the same cluster if the cluster contains at least two hubs.
4. Sum the product of flow volumes W_{ij} and the shortest-path distances over all pairs of nodes i and j to obtain a lower bound of the minimum capacities.

5. The lower bound can be tightened by estimating the increase in cost if a particular node i in a cluster containing one or more hubs is in fact not a hub.

Branching simply occurs by sub-dividing a cluster and enumerating all possible node allocations between the sub-divisions of the cluster. Note that any lower bound for the multiple- allocation problem is also a lower bound for the single-allocation problem. In order to obtain a feasible single-allocation solution, further branching may be required to uniquely allocate a non-hub node to a hub node.

3.6.4 Clustering heuristics

Klincewicz (1991) used the clustering heuristics methodology as one of the methods for choosing hubs in the facility-location problem. The area was divided into clusters and the different airports were given indexes in terms of probabilities, using the principle that the airport in a cluster that is most suitable as a hub would be the airport with the shortest node-hub distances and the highest passenger demand. Matrices with data showing distances and passenger numbers to and from all the airports within the clusters are collected. Probability indexes are applied to each of these matrices such that for each origin airport:

- The destination node with the shortest distance from the origin will have the highest index of 1 for the distance matrix.
- The destination node to which the largest number of passengers from the origin node is flying has the highest index of 1 for the passenger matrix.

The indexes are totalled up for each node within the cluster that is a favourable destination as a hub in terms of:

- the node with the highest total index being the one with the shortest distance to all the nodes in the cluster
- the node with the highest total index being the one with the highest passenger flow to all the nodes in the cluster
- the node with the highest total index being most favourable in terms of both distances and passenger flow.

The hub with the highest total index will then be chosen as the most probable hub. The method of clustering was shown to have the advantage of narrowing down a search from a large number of nodes over a whole network to fewer nodes within a cluster, making it a more effective way of optimising the movement of flow.

3.6.5 Direct Vs non-stop services

One of the disadvantages of hubbing is the inconvenience of not having direct flights from one node to another. Aykin (1995) suggested that in a bid to improve service in air passenger transportation, more convenient non-stop flights can also be offered by the airlines between some non-hub cities. For each route, a decision regarding the service type is made such that flow between two cities is either shipped non-stop (direct shipping with no hub stop) or shipped through hub(s) (one or more hub stop).

Even though single-hub assignment has the advantages of network simplicity and possible higher facility utilisation, this may not be acceptable because of the operational restrictions it imposes on the system. Passengers may choose services that are more convenient rather than making one or more hub stops or having to take long detours every time they fly.

Aykin (1990) considered the discrete hub-location and routing problem with either the non-stop, the one-hub-stop or the two-hub-stop services for a hub network in which the hubs had already been located and their capacity was known. This procedure can be used to allow for flexibility especially if capacity problems do occur on some routes, causing hubs to reach their capacity for origins of high passenger demand. This procedure in turn allows for the hub-network to compete favourably with the traditional passenger airlines for these routes.

Aykin (1995) formulated the problem for location-allocation in which the hub locations or the service types are known. The problem is decomposed into a number of shortest-path problems involving service-type decisions if the hub locations are available. And if the service types are known, then the problem is reduced to a multi-facility location problem.

3.7 Summary

This summary is derived from the literature reviewed above on designing an H&S network, with particular application to the African region. The main aim of creating an H&S network is to minimise air transport costs over the vast African continent, which has sparse passenger demand. The network will focus on trying to consolidate passenger demand along the routes while transporting passengers from their origins to their destinations through hubs, which is one of the major benefits of the H&S network. The limitation of the design methodology is that the method for network design uses values that are manually input into a network cost equation from the cost model, thus the methodologies of heuristics, Tabu Search, Genetic Algorithms and linear programming cannot be used to find the optimum network. In order to use the above-stated methodologies, automation of the cost model would be necessary, yet the advantage of this cost model is that it recalculates the most cost-effective option for each route in the network, in terms of operational and service parameters.

3.7.1 Hub location

The hub-location procedure will be taken as the ρ -hub median problem, where a fixed number of hubs (ρ) are chosen from n nodes, which are the airport locations. The hub-location problem will be solved using various methodologies with cost justifications. Africa faces the dilemma of not having many airports with the capacity and infrastructure for hubbing in terms of runways, gates and slots because of the low passenger demand and the number of flights operated. For the purpose of this study, the present airport capacity in terms of demand and infrastructure is ignored since the majority of African airports lack the proper infrastructure. As a first-cut analysis for Africa, the possible hub airports will be chosen based on the most suitable geographic location that would reduce the total network costs.

3.7.2 Node allocation

The node allocation will be solved as the Uncapacitated Single-Allocation ρ - Hub Median Problem, which implies that each node will be assigned to only one hub; this is done to limit the complexity of network cost calculations and operations. Furthermore, all nodes have to be routed via one hub, namely the hub with the closest distance to the node, to gain the benefits of flying short routes which use smaller aircraft which are cheap to operate.

3.7.3 Hub-and-spoke network

The total cost for the network is defined as the total cost of moving passengers from their origins to their destinations. The main approach in the literature to minimising the costs of transporting flow from origin to destination in an H&S airline network is that established by operational researchers who calculate the cheapest hub-location options that will lower network costs.

In this study, lowering the costs of the hub network will be carried out in two ways:

1. The total network costs will be minimised using the linear quadratic equation developed by O’Kelly et al. (1986) which was revised as shown in Equation 11 by Klinecicz (1991) to yield an equation that can be applied to larger, more complex networks so that the solutions can be evaluated more efficiently. The first part of Equation 11 calculates the node-hub costs, while the second part calculates the hub-hub costs:

$$f(\mathbf{x}) = \sum_i \sum_k X_{ik} C_{ik} (O_i + D_i) + \sum_i \sum_k X_{ik} \sum_k \sum_m X_{km} \alpha C_{km} W_{km} \quad \text{Equation 11}$$

2. The lowest cost per passenger and the passenger numbers used in Equation 11 for each route in the network will be derived from the cost model developed by Ssamula (2004). The cost model calculates the operating costs incurred by flying along a specified route, and the database for this model contains Africa-specific data. The costs used are calculated by selecting the aircraft (chosen from 11 different aircraft types of varying capacity) most commonly used in Africa that produces the lowest operating costs for the route. A full description of the model is given in Chapter 4.