Strategic Planning of Large-scale, Multimodal and Time-definite Networks for Overnight Express Delivery Services

M. Sc. Yida XUE
Technische Universität Darmstadt – Germany
Section Supply Chain and Network Management
Department of Laws and Economics
D17

Strategic Planning of Large-scale, Multimodal and Time-definite Networks for Overnight Express Delivery Services

M. Sc. Yida XUE

born in Wuxi, China

Supervisor: Prof. Dr. Dr. h.c. Hans-Christian Pfohl

Co-Supervisor: Prof. Dr. Malte Fliedner

Submission date: 30th January, 2013

Examination date: 7th May, 2013

Technische Universität Darmstadt

Section Supply Chain and Network Management

Darmstadt, Germany

D17

Strategic Planning of Large-scale, Multimodal and Time-definite Networks for Overnight Express Delivery Services

Vom Fachbereich Rechts- und Wirtschaftswissenschaften der
Technischen Universität Darmstadt genehmigte

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum politicarum (Dr. rer. pol.)
vorgelegt von

M. Sc. Yida XUE aus Wuxi, China

Referent: Prof. Dr. Dr. h.c. Hans-Christian Pfohl

Korreferent: Prof. Dr. Malte Fliedner

Einreichungstermin: 30. Januar 2013

Prüfungstermin: 7. Mai 2013

Darmstadt 2013

D17

Acknowledgements

This dissertation is the results of my research work as a PhD candidate in the Faculty of Laws and Economics at Darmstadt University of Technology, Germany. My work would not have been possible without the guidance, insights, help, support and encouragement from many people around me.

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Dr. Dr. h.c. Hans-Christian Pfohl. He gave me the opportunity to do my research work in TU Darmstadt and become a member of his active research team. I also own him thanks for his guidance and encouragement for me to learn the German language and do my research partially in German.

I am thankful and indebted to my thesis mentor Professor Dr. Malte Fliedner for his enthusiasm, input, comments and guidance for my research work, especially in the field of modeling and algorithms. I am very thankful for his luminous suggestions that had challenged me to think about my research more critically, his careful corrections on my work and the long but very pleasant discussions I had with him.

I want express my gratitude toward Professor Dr. Jiazhen Huo. I want to thank him for his constant support and encouragement to me since my graduate study in Tongji University and his immense efforts to help me work in the enterprise project and identify this research topic.

It is also my great pleasure to thank the members of the project committee for their pleasant cooperation, suggestions on research orientation and sharing of their source data. Special thanks to Guofeng Han for his hard work on coding and data compiling.

I am extremely grateful to Professor Dr. Dr. h.c. Wolfgang Domschke for the documents and scientific instructions he gave me during my preliminary research on this topic.

I am also thankful to Professor Dr. Thaddeus Kim Teck Sim for his generosity to share research data with me.

It is my great honor to work with my colleagues in the section of Supply Chain and Network Management. I would like to thank Dr. Nguyen Thi Van Ha, Dr. Jian Cui, Dr. Xin Shen, Dipl.-Ing. Markus Ehrenhöfer, Dipl.-Ing. David Thomas, Dipl.-Ing. Christian Zuber and M.Sc Yanqiang Ma for their help and support during my time in Germany.

Finally, I express my great gratitude to my parents for their endless support and love.

Yida Xue

Darmstadt, 30th January, 2013

Contents

Abstract

The rising demand of express delivery service (EDS) and fierce market competition motivate EDS providers to improve service quality by modifying current networks. This project-based dissertation focuses on strategic planning of a large-scale, multi-modal and time-definite EDS network for a top nationwide EDS provider in China, based on its current network.

An air-ground Hub-and-Spoke (H/S) network with a fully interconnected/star shaped structure was established to provide trans-city overnight EDS among relatively developed cities in China. The corresponding models are a combination of the hub location problem with fixed cost and the hub set covering problem. The objective function is to minimize the sum of the hub-location fixed cost and transportation cost under the constraints that all demand nodes are covered by their "home" hub. First, the basic model with linear air cost was proposed. Next, the basic model was extended to include air service selection decisions (or aircraft fleet ownership decisions) under the consideration of a cost select function for the backbone air service. Finally, two extension models were studied, one to obtain the optimal aircraft fleet composition (*Ext.1*) and the other under the constraints of current aircraft fleet composition (*Ext.2*).

Due to the large scale of project instances, hybrid genetic algorithms (GAs) were applied to get desirable solutions in an acceptable time period, but without the guarantee of finding optimal solutions. In particular, the overall problem includes three kinds of decisions: 1) hub location decisions, 2) demand allocation decisions and 3) air service selection decisions. A specific algorithm was proposed for each kind of decision, namely, GAs, local search heuristics and integer programming, respectively. These three algorithms were invoked hierarchically and iteratively to solve the original problem. 5 improvement techniques were proposed to different procedures of the original algorithms in order to improve the performance of the algorithms.

Computational tests were conducted to evaluate the performance of the proposed algorithms in terms of computational time and solution quality. Tests under small-scale instances with CAB data sets were conducted to evaluate the overall performance of the proposed algorithm by comparing the solutions with the optimal solutions generated by CPLEX. Tests under large-scale instances with AP data sets and project data sets were conducted to evaluate the performance of the proposed improvement techniques. Since neither the optimal solutions nor solutions by other algorithms under large-scale instances were available to serve as benchmarks, the performance of the tailored algorithms and that of the un-tailored simple GAs was compared. Information about the stability of the algorithms with values of the coefficient of variation (CV) and the reliability of the results with T-tests was also provided.

The models and the tailored GAs were applied to real-life instances of the project. This study introduces how the input data were collected and modified and how to deal with pertinent problems. By analyzing and comparing the basic solutions of Ext.1 and Ext.2, the study not only reveals some important features of the network, but also arrives at some general conclusions and provided a dynamic aircraft fleet update strategy to guide the implementation of the project. Finally, scenario planning was executed to help decision-makers balance between costs and corresponding decision risks by identifying critical uncontrollable and controllable factors.

Zusamenfassung

Die steigende Nachfrage nach Express-Delivery-Service (EDS) sowie starker Marktwettbewerb veranlassen die EDS-Dienstleister, ihre Service-Qualität zu verbessern, indem sie ihre derzeitigen Netzwerke modifizieren. Gegenstand der vorliegenden, projektbasierten Dissertation ist die strategische Planung eines großmaßstäblichen, multimodalen und zeitbestimmten EDS-Netzwerks für einen Top-Anbieter für landesweiten EDS in China, basierend auf seinem derzeitigen Netzwerk.

Es wird ein Luft-Boden-Netzwerk eingeführt, mit einer vollständig vernetzten, sternförmigen Struktur, um interstädtische Über-Nacht Lieferung zwischen entwickelten Städten Chinas anzubieten. Die entsprechenden mathematischen Modelle kombinieren das Problem der geographischen Festlegung der Hubs und der damit einhergehenden Fixkosten mit dem Problem des Set-Covering. Dabei wird die Summe der Fixtenkosten der Hubs und der Transportkosten minimiert, mit der Bedingung, dass alle Knotenpunkte des Netzwerks (engl. demand nodes) von ihren „Heimat-Hubs" versorgt werden. Zunächst wird das grundsätzliche Modell mit linearen Flugkosten vorgestellt. Dieses wird sodann um die zusätzliche Berücksichtigung der Entscheidungsprobleme der Auswahl einer Fluggesellschaft bzw. Anschaffung einer eigenen Flugzeugflotte erweitert. Dabei wird eine Kostenwahlfunktion für das Hubnetzwerk zugrundegelegt. Es werden zwei erweiterte Modelle untersucht, eines mit optimaler Flugzeugflottenzusammenstellung (*Ext.1*) und das andere unter der Bedingung der Beibehaltung der aktuellen Flottenzusammenstellung (*Ext.2*).

Wegen der Ausmaße der zu untersuchenden Fälle wurde zur Lösung der gestellten Aufgabe auf hybride GAs zurückgegriffen, um gute Lösungen in vertretbarer Zeit zu bekommen, aber ohne die Garantie, die optimalen Lösungen zu finden. Im Einzelen hat das Problem drei Entscheidungsbereiche: Entscheidungen zur Standortwahl für die Hubs, zur Zuordnung der Nachfrage und schließlich zur Auswahl des Luftfrachtdienstes. Für jeden Entscheidungsbereich wird ein spezifischer Algorithmus vorgeschlagen, nämlich GAs, lokale Such-Heuristiken bzw. binäre Programmierung. Diese drei Algorithmen durchlaufen sukzessive einen vorgegebenen hierarchischen Ablaufplan, um das ursprüngliche Problem zu lösen. Um die Performance der Algorithmen zu verbessern, werden 5 Verbesserungstechniken entwickelt, die auf verschiedene Rechenschritte des ursprünglichen Algorithmus angewendet werden.

Rechentests werden durchgeführt, um die Performance hinsichtlich der Rechenzeit und Effektivität der Algorithmen zu beurteilen. Tests in kleinem Maßstab mit Datenmaterial von CAB sollen die Gesamtperformance des entwickelten Algorithmus durch Vergleich der vorgeschlagenen Lösungen mit den durch CPLEX erzielten optimalen Lösungen evaluieren. Tests unter großmaßstäblichen Bedingungen mit dem AP-Datenmaterial und Projektdaten dienen der Evaluierung der Performance der entwickelten Verbesserungstechniken. Da weder die optimalen Lösungen noch Lösungen von anderen Algorithmen für Fälle mit großem Ausmaß zur Verfügung stehen, um als Benchmark zu dienen, wird die Performance der angepassten Algorithmen mit derjenigen der einfachen GAs verglichen; mit Varianzen und Koeffizienten, um Informationen über die Stabilität der Algorithmen zu bekommen, bzw. mit *T*-Tests, um die Zuverlassigkeit der Ergebnisse zu überprüfen.

Die Modelle und angepassten GAs werden auf reale Fallbeispiele im Rahmen des Projekts angewendet. Es wird gezeigt, wie Inputdaten gesammelt und modifiziert werden und wie auftretenden Problemen begegnet werden kann. Durch Analyse und Vergleich der Resultate von *Ext. 1* und *Ext.2* werden nicht nur einige

wichtige Eigenschaften des Netzwerks aufgezeigt, sondern es werden auch einige allgemeine Schlussfolgerungen gezogen sowie eine dynamische aktualisierte Strategie für die Flugzeugflotte erarbeitet. Schließlich soll Entscheidern durch Entwicklung von Szenarien geholfen werden, kritische Faktoren zu erkennen, Unsicherheiten zu beseitigen und zwischen Kosten, Risiken und Nutzen abzuwägen.

List of Figures

List of Tables

List of Abbreviations

AP data	Australia Post data
AI	artificial intelligence
B&B	branch-and-bound
B2B	business-to-business
B2C	business-to-customer
CA	cluster analysis
CAB	Civil Aeronautics Board
CAE	China Air Express
CGAs	Constructive GAs
CHLP	capacitated hub location problem
CPLNFP	concave piecewise linear network flow problem
COPs	combinatorial optimization problems
CRE	China Railway Express
CV	coefficient of variation
C2C	customer-to-customer
EA	evolutionary algorithms
EDS	express delivery service
EDSI	express delivery service industry
EOP	economy of scope
EOS	economic of scale
ESP	express service provider
FLP	facility location problem
FLS	full local search
GAs	genetic algorithms
GDP	gross domestic product
GLS	genetic local search
GRASP	greedy randomized adaptive search procedure
HCFP	hub covering flow problem
HLP	hub location problem
HLRP	hub location-routing problem
H/S	hub-and-spoke
ILS	iterated local search
IPO	initial public offering
JIT	just-in-time
LAP	location-allocation problem
LATN	Local Area Transportation Network
LOS	level-of-service
LRP	location-routing problem
LS	local search
LSP	logistic service provider
LTL	less-than-truck

MAs	Memetic Algorithms
MALT	multi-start alternate algorithm
MCNFP	Minimum Concave-Cost Network Flow problem
MIP	mixed-integer programming
MS	management science
MTSP	the multiple traveling salesman problems
NP-hard	non polynomial hard
O-D	origin-destination
OR	operations research
PCA	principal component analysis
PLS	partial local search
PO	post office
RCL	restricted candidate list
SA	Simulated Annealing
SGAs	simple genetic algorithms
SNDP	service network design problem
SSP	shipment service provider
TS	Tabu Search
UHLP-S	uncapacitated hub location problem with single allocation
USITC	The United States International Trade Commission
VNS	variable neighborhood search
VRP	vehicle routing problem
WTO	The World Trade Organization (WTO)

1. Research background and preparation work

1.1. Introduction of express delivery service (EDS)

This chapter is a preparation for the research by introducing research background, the project and prior work for our research. In Sec.1.1 we generally introduce EDS and its development. In Sec.1.2 we introduce the project sponsor and its motivation, and put our research topic in the project framework. In Sec.1.3 we introduce some preparation work for our research. Finally, in Sec.1.4 we propose the outline of the dissertation.

1.1.1. Definition of EDS

First of all, we would like to enumerate several well-acknowledged definitions of EDS to provider readers a general idea about what is EDS.

The World Trade Organization (WTO) defines EDS as “a single or multi-model transportation service offered by non-postal firms rather than national post offices, including collection, transport and delivery of mails and packages both internationally and nationally. This service can be accomplished by self-owned or public vehicles.”[1]

The United States International Trade Commission (USITC) defines EDS as: “the expedited collection, transport and delivery of documents, printed matters, parcels or other goods, while tracking their location and maintaining control over them throughout the whole service process; relevant services include such as customs facilitation and logistics services; accessary and value-added services include, for example, collection from a location designated by the consignor, release upon signature, guaranteed specified delivery time, and delivery confirmation.”[2]

According to the National Bureau of Statistics of China and the State Post Bureau of The People's Republic of China, “EDS is the service that a carrier transports or deliveries goods using the fastest transportation mode to the designated destination within specified time, keeps transportation and delivery under control and provides real-time information.” [3]

The Administrative Measures for Express Market defines EDS as “a service that involves expedited collection, transport and distribution of individually sealed/packed mails and parcels to consignee or designated place within specified time and release upon signature.” [4]

[1] This definition is contained in CPC7512. Definition and other related rules about the topic EDS, please refer to official website of WTO http://www.wto.org/.

[2] See the United States International Trade Commission (USITC) (2004), p.1-1. Available at website: http://www.usitc.gov/publications/332/pub3678.pdf (access on 19.01.2013).

[3] See Research on Development of Express Market in China (2007), p.2.

[4] See the Administrative Measures for Express Market issued by Ministry of Transportation of the People’s Republic of China in July, 21, 2008. Also available at website: http://www.gov.cn/flfg/2008-07/30/content_1059671.htm (access on 19.01.2013).

In this dissertation, EDS refers to nationwide mail and package delivery service within specified time with multi-modal transportation networks.

1.1.2. Development of the EDS industry (EDSI)

With the rapid development of international trades in the 1960s and 1970s, the delivery speed and service quality of regular postal services could not keep up with the pace of the modern economic development. In 1969, three budding entrepreneurs - Adrian Dalsey, Larry Hillblom and Robert Lynn- founded DHL in San Francisco, personally shipping papers by airplane from San Francisco to Honolulu[5]. It's the embryonic form of the modern EDS.

At that time trading and banking industries required expedited information delivery in order to achieve higher efficiency. To fulfill this requirement, EDS boomed robustly all over the world, thanks to its swiftness and security. In the 1990s EDSI has taken shape in developed countries and/or regions, such as the United States, Japan and Europe. In recent years, EDSI has also thrived in developing countries, owing to the local social and economic development. In 2008 EDSI made a direct contribution of USD 80 billion to the world GDP (equivalent to the nationwide ship manufacturing industry) and also provided 1.3 million jobs in the world directly, in addition to another 2.75 million jobs indirectly[6].

Ten years after the first express company in the world came into shape, this new service philosophy and operational model was introduced to China. In June 1979 the first Chinese express delivery company was founded together by SINOTRANS from China and OSC from Japan[7]. China Post started to offer international and national EDS in 1980 and 1984 respectively. In October 1985, China Express Service Company was established[8]. It had taken a dominant position in the Chinese EDS market till the first half of the 1990s, when it was the exclusive nationwide EDS provider.

After the year 1992, the development of export-oriented economy in delta regions of Yangtze River and Pearl River motivated the growth of private economy in those regions. Enterprises in those regions participated more intensively in the international division of labor. Postal service could subsequently no longer meet the increasing requirement of fast, convenient, reliable but less expensive door-to-door delivery service of documents, samples and catalogues. Private and nongovernmental enterprises emerged in response to this requirement. Some firms even began to use air-ground network in order to offer time-definite package delivery service.

Later four magnates in the international EDSI entered into Chinese market consecutively. In Dec. 1986, DHL-SINOTRANS was founded in Beijing. A joint venture was established by TNT Express and SINOTRANS Group. FedEx started its own direct flight to China and built its own logistic infrastructure in China quickly after that. Finally in Apr. 2001, UPS also expanded its business in China.[9]

---

[5] See official website of DHL: http://www.dhl.de/de/ueber-uns/unternehmensportrait.html (access on 19.01.2013).

[6] See Oxford Economics (2009), p. 7.

[7] See Wang/ Liu (2000), pp. 36-37.

[8] See website: http://www.ems.com.cn/aboutus/fa_zhan_li_cheng.html (access on 19.01.2013).

[9] See Research on Development of Express Market in China (2009), pp.3-5.

After China's accession to the WTO in 2001, China has promised that from Dec.11th, 2004 foreign-invested logistic enterprises can establish exclusively-founded enterprises in China and that from Dec. 18th, 2004 internal EDS market in China is officially opened to these exclusively-founded enterprises[10]. From that time major international EDS providers have intensively made investment, promoted market and enhanced service quality in China. China Post's market share of internal EDS has decreased annually with an average rate of 4% since 2005, while the four international magnates, namely FedEx, UPS, DHL and TNT, keep the growth of annual sales volume for more than 30%.[11]

1.2. Project introduction

1.2.1. Introduction of Company A

Company A, the sponsor of this project, is a state-owned enterprise in China. Its internal business covers more than 2000 cities in 31 provinces, autonomous regions and nationally administered municipalities all over China with more than 30 thousand service points and sales agencies.

To be more specific, internal EDS by Company A is based on a transportation network composed of a letter-and-mail delivery network, express-exclusive flights and express handling centers. Its current air transportation network is supported by a self-owned aircraft fleet, a large number of commercial flights and 45 air transportation handling centers. In July 2007 a new air hub was established in Nanjing. Fig.1-1 shows the business volume of internal EDS by Company A in recent years.

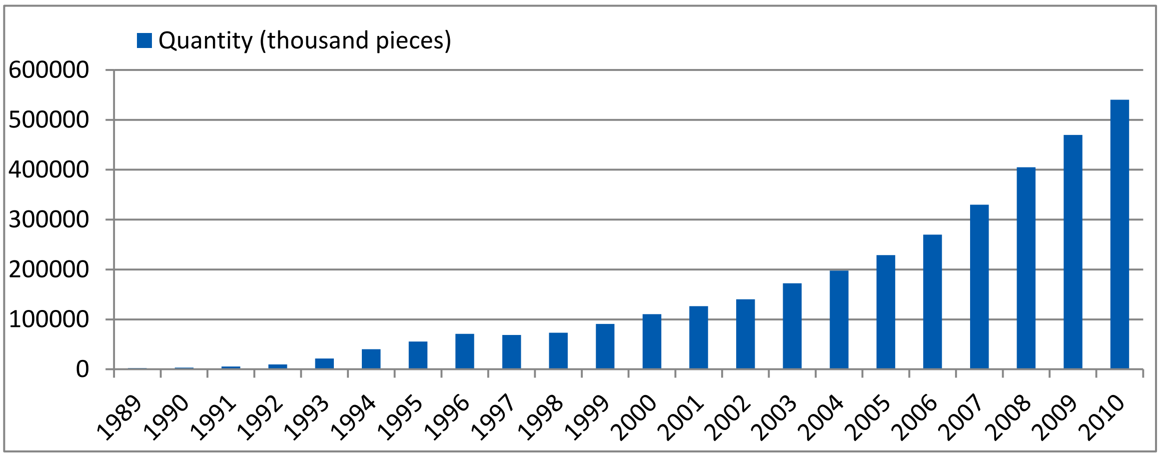

Figure 1-1: Business volume of Company A

(Source: based on company's annual report)

1.2.2. Project motivation

1.2.2.1 Policy factors

Policy pressure mainly comes from China's commitment to the WTO. Foreign Market Access and National Treatment Clause are the two of the most important agreements among the General Agreement on Trade in

[10] See official website of WTO: http://www.wto.org/ (access on 19.01.2013).

[11] See China Logistic Yearbook (2008), p.168.

Services (GATS) of the WTO.[12] Form unilateral regulations and constraints that foreign firms had to follow have been eliminated according to China's commitments to the WTO, such as the constraints on share ratios of logistics and freight forwarding companies, limited trucking licenses and limited integrated services in China. Moreover, according to the National Treatment Clause in GATS, local EDS providers cannot enjoy preferential policies, such as tax reduction, national subsidy, priority on highway, exemption from traffic control in metropolis and rapid channel for customs clearance. Otherwise the same preferential policies should also be applicable to foreign-invested service providers.

1.2.2.2 Rise in demand

It was estimated that the EDSI generated sales revenue (i.e. turnover) of USD 175 billion globally in 2008. Stripping out the inflation effect, the sales revenue of EDSI is estimated to have risen by over 20% since 2003 at an average annual rate of 4%, slightly faster than the growth rate of the world economy (Fig.1-2 shows the increase percentage compared with 2003). According to the estimation by economists from Oxford University, the contribution of EDSI to the world GDP will reach USD 1350 billion in 2013[13].

Fig.1-3 displays the fast growth of business volume of the EDS in China. Market size of EDS in China is strongly correlated to its GDP. Specifically, 1% increase in the GDP corresponds to 3% increase in the EDS market[14]. The hysteretic development of second-tier cites injects extra energy to the rise of EDS demand. Hence, the growth of EDS market in China will maintain over 25% in the following years. It is also estimated that the whole EDS market size in China will reach 163.2 billion USD in 2020, 82.5 billion of which comes from nationwide EDS[15].

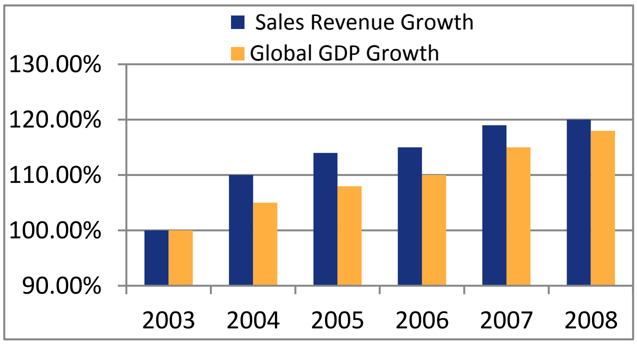

Figure 1-2: Sales revenue of global EDSI versus global GDP(compared with 2003)

(source : based on Oxford Economics(2009), p.5)

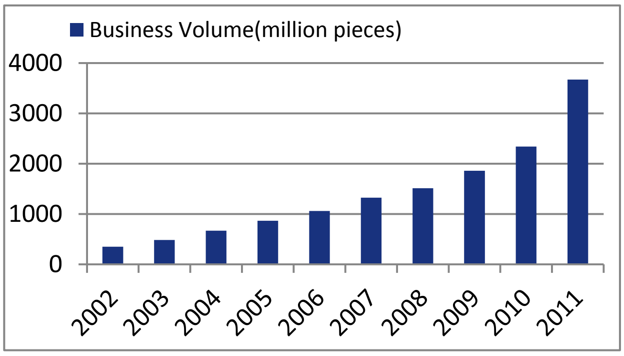

Figure 1-3：Business volume of EDSI in China [16]

The fast expansion of the EDS market size in China results from the development of the high-tech, finance and service industries who are the major users of EDS (see Fig.1-4). Besides, two obvious trends also stimulate the

[12] See http://www.wto.org/.

[13] See Oxford (2009), p.37.

[14] See Research on Development of Express Market in China (2009), p17.

[15] See Research on Development of Express Market in China (2009), p26.

[16] Data from 2002-2006 comes from Zhang/Zhao (2006), p34; Data from 2007-2011 comes from official website of China Post: http://www.spb.gov.cn/folder7/folder31/index_2.html (access on 19.01.2013).

demand of EDS in China.

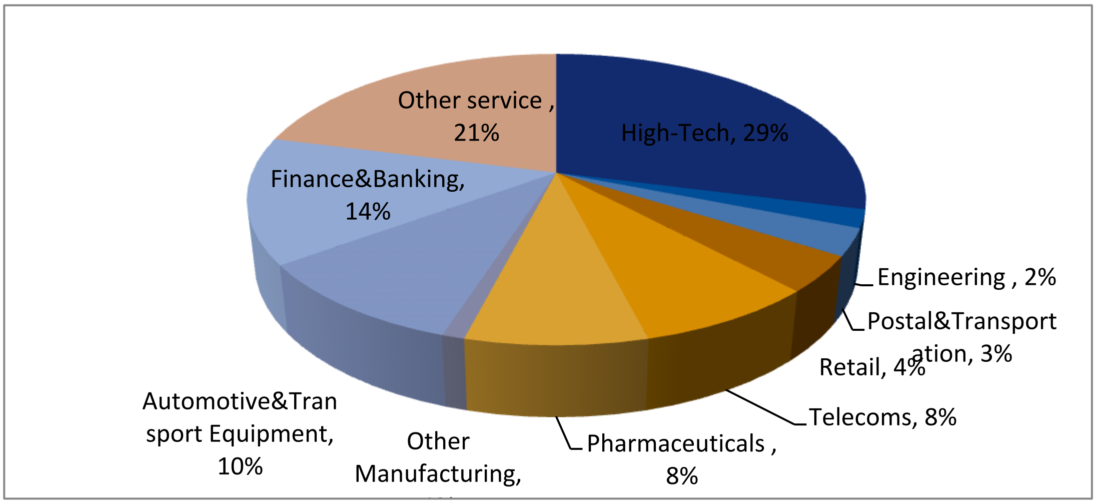

Figure 1-4: Major clients of EDSI

(Source: based on Oxford Economics (2009), p.8)

(1)Boom of E-Commerce

Although EDS providers already partner with traditional catalog firms, C2C (consumer-to-consumer), B2B (business-to-business) and B2C (business-to-consumer) transactions over the Internet push up the EDS demand. Private consumers in China begin to buy electric devices, toys, clothes and even furniture online. Moreover, enterprises purchase accessories and nonconventional raw material through online B2B platforms, such as Alibaba and Conghui360[17].

Following the development of E-Commerce, EDS in China has witnessed a rapid growth. Close cooperation between EDS providers and E-Commerce platforms, such as Alibaba and Taobao[18], pushed C2C and B2C online sales volume to USD 21.3 billion in 2008, taking up nearly 1% of the total sales volume of the whole society. Statistical data show that E-Commerce in China created more than 500 million pieces of EDS order in 2008, taking up almost half of all business volume of EDSI. Despite of the financial crisis in 2008, mid-size and large-size EDS providers in China achieved sales volume of about USD 6.65 billion (19.2% more than the year before) and accomplished 1.5 billion pieces of order[19].

(2) Prevalence of Just-in-time (JIT) and outsourcing strategy

Demand for EDS is increasing rapidly as a result of the prevalence of JIT and outsourcing strategies among Chinese manufacturing enterprises. Compared with the high cost of large inventory, the cheap price of the EDS attracts manufacturers to use smaller but more frequent shipments for intermediate and final products.

1.2.2.3 Current unreasonable service network

[17] See their official websites http://www.alibaba.com/ and http://www.hc360.com/ (both access on 19.01.2013).

[18] See its official website http://www.taobao.com/ (access on 19.01.2013).

[19] See Research on Development of Express Market in China, (2009) p.23.

As we have mentioned, Company A currently shares its service network with a letter-and-mail delivery service provider, especially the ground transportation network. However, EDS and mail service are different essentially. Letter-and-mail service is widely offered to the whole society with relatively low price, while EDS is oriented to special requirements or even customized requirements, for instance, desk-to-desk and time-definite delivery. Price of EDS is subject to the market discipline and depends on service level and also supply-demand relations. Therefore, networks for letter-and-mail service should be widely spread, while networks for EDS should pay more attention to time efficiency and operational flexibility. A new service network with high air gateway density, seamless ground service and short road transportation is in demand.

The above-mentioned factors motivate Company A to carry out the project and redesign its service network to meet the market challenge, allocate limited resources more efficiently and satisfy customer requirement, so that Company A can maintain its advantage and gain a larger market share.

1.2.3. Project framework and research boundary

The framework of the whole project is displayed in Fig.1-5.

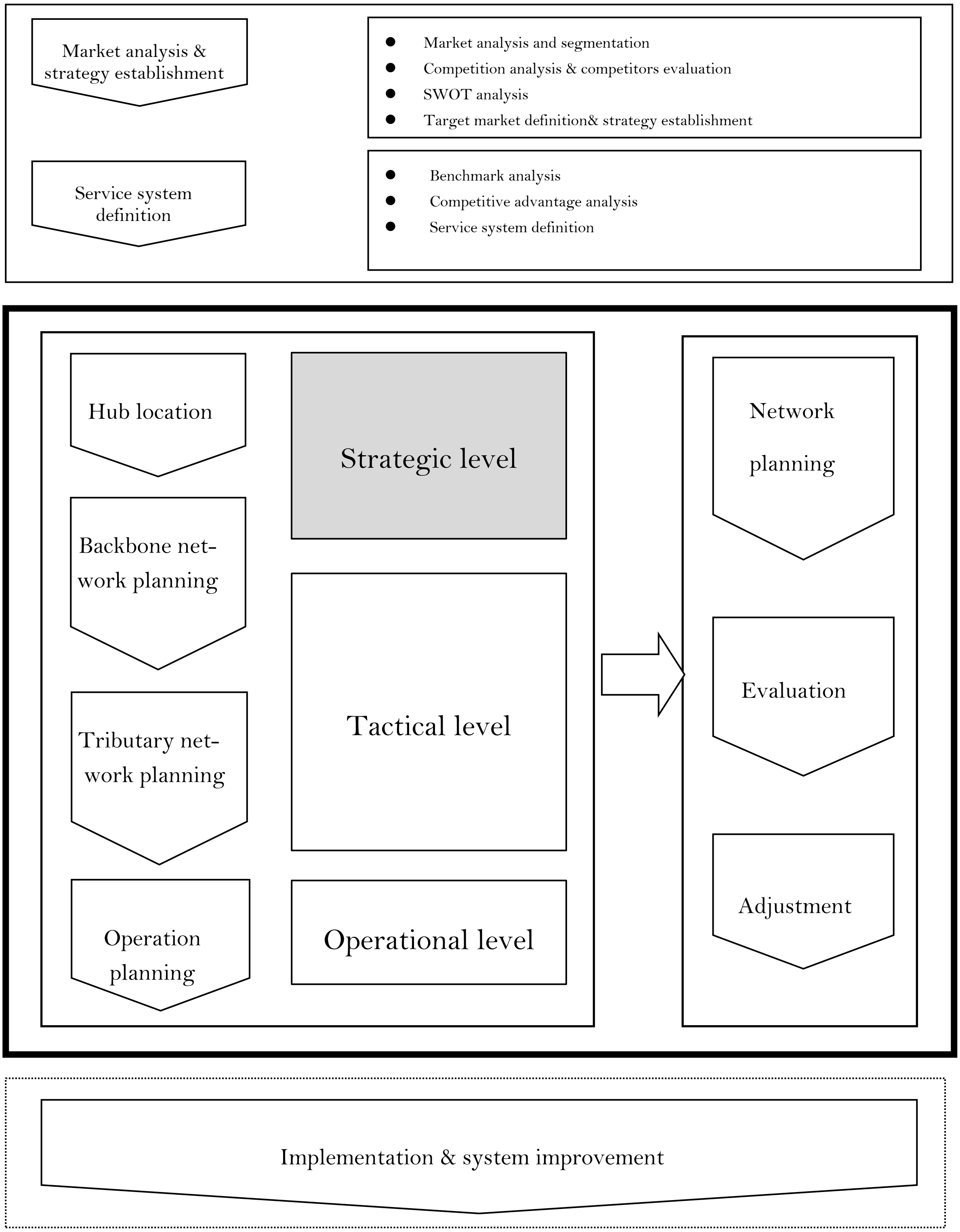

Figure 1-5: Project framework

The project has three phases. The first phase described in the top box includes market strategy establishment and service system definition. The last phase of the project is the implementation of the planning work and system improvement. The main body of the project is the second phase: EDS system design, which includes three planning levels. Strategic planning considers those decisions with long-term impacts on the network, including hub location and aircraft acquisition decisions. Tactical planning involves backbone network and tributary network planning under the constraints of the strategic decisions. Operational issues, such as routing and scheduling, are tackled at operational planning level.

This dissertation is dedicated to the strategic planning of the network (the shadowed part in Fig.1-5), including hub location, demand allocation and air service selection decisions. The input of our study is demand nodes and potential hubs sets, definition of services, demand volume and relevant cost information.

1.3. Preparation work

What kind of service Company A is going to provide determines the configuration and features of the transportation network it needs. As preparation work of our research, this section briefly introduces how we position the target service with marketing tools. Input data collection and modification will be introduced in Chap.6.

1.3.1. Market segmentation

Philip Kotler et al argued that market segmentation, target market selection and service positioning constitute the prerequisite steps in designing a successful marketing strategy.[20] These steps guide the enterprise to focus its efforts on the right customers. We follow these three steps to position the target services.

The concept of "market segmentation" was introduced by Wendell Smith in 1956[21]. Although many definitions of market segmentation have already been proposed, the original one proposed by Smith still keeps its value: "Market segmentation involves viewing a heterogeneous market as a number of smaller homogenous markets, in response to different preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants."[22] Market segmentation helps organizations to identify market opportunities, improve the allocation of resources, develop a competitive market position and ultimately lead to more satisfied customers[23].

We divide the EDS market in China according to the geographical scope and service quality (see Fig.1-6). In the dimension of geographical scope, the EDS market in China has three market segments- international, nationwide trans-city, and intra-city EDS. In the dimension of service quality, high, median and low service levels are available in each market segment.

International EDS is a capital and technology intensive business that also yields the highest profit. The requirement of a wide-spread service network all over the world and advanced information systems excludes most service providers outside of this high-end market segment. More than 80% of the international EDS market in China is currently occupied by the four international express tycoons-DHL, TNT, UPS and FedEx.

[20] See Chernev/Kotler (2008), pp. 1-2.

[21] See Smith (1956), p3.

[22] See Smith (1956), p.3.

[23] See Wind (1978), p.317.

Since nationwide trans-city EDS requires a service network all over China, this market segment is shared by few large enterprises. At the same time, 80% of the business volume centralizes in and between the delta regions of Pearl River and Yangtz River, also Bohai Bay economical region. Meanwhile, some medium-size non-governmental and private enterprises also provide trans-city but not totally nationwide EDS, focusing on one of these regional markets or connecting several major cities between areas.

Intra-city EDS is a labor-intensive service that experiences the fastest growth in last few years. Without high technology and large investment on vehicles and distribution centers, many small nongovernmental and private enterprises have entered into this fiercely competitive market, taking up nearly 80% of the market share.[24] Unbeatable operational flexibility and low labor cost enable these enterprises to offer fast desk-to-desk intra-city EDS with quite low price. Fig.1-6 illustrates the major participants and their dominant positions in different market segments.

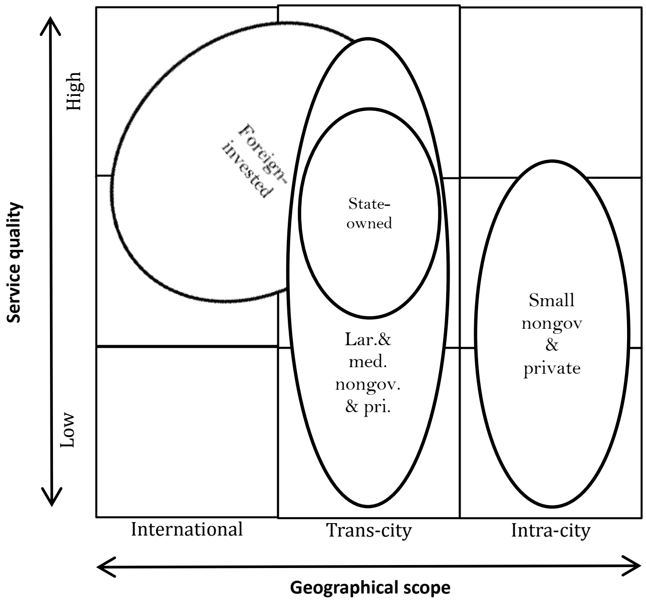

Figure 1-6: Market participants and their positions in market segments

1.3.2. Target market selection

Target market selection requires comprehensive understanding of the participants on the market. Different participants have corresponding strengths in different market segments. Every enterprise should understand the opportunity and risk outside, as well as the advantage and weakness of itself. It should make full use of its advantage and catch the opportunity, while eliminating its weakness and avoiding risk[25].

[24] See The First Statistic Survey on Express Delivery Service (2007), available online: http://www.spb.gov.cn/folder7/folder31/2007/07/2007-07-24173.html (access on 19.01.2013).

[25] See Pfohl (2004), p.82.

Although Company A is still the leader in the nationwide trans-city EDS market, it has gradually lost its market share, since it was reluctant to reform its inefficient service system and bureaucratic management mechanism. Most customers could neither accept the high price nor bear the poor service. Meanwhile, some nongovernmental and private enterprises expanded their business to nationwide scope. Some large state-owned enterprises began to reorganize and work together with local ground carriers to offer nationwide door-to-door EDS based on their resource advantages, taking China Railway Express (CRE) and China Air Express (CAE) as examples. Subject to the Chinese Postal Law issued in 1986, foreign-invested EDS providers were not allowed to offer nationwide EDS in China until China's accession into the WTO. That means they were only allowed to offer inbound and outbound delivery service that was directly triggered by their international service. For this reason, at that time they could enter the Chinese internal EDS market only by cooperating with Chinese local companies, taking SINOTRANS as an example. However, this situation has changed after China's entry into the WTO. After Dec. 18th, 2004 when China's internal EDS market was first opened to the outside world, foreign-invested enterprises began to expand their service network in China rapidly, from major cities to second-tier cities.

Currently, three major market participants, namely state-owned enterprises, nongovernmental and private enterprises and foreign-invested enterprises, share the nationwide trans-city EDS market in China. According to the statistics by the National Post Bureau, in 2006 state-owned, nongovernmental and private and foreign-invested enterprises contributed 49.5%, 17.5% and 33% of the total sales volume of the trans-city EDS in China, respectively (see Fig.1-7), and 58.4%, 27% and 14.6% of the total business volume(see Fig.1-8)[26]. Following after this analysis, we look deeper into the three players in the nationwide trans-city EDS market. Please refer to Tab.1-1 for the detailed comparison of their service networks.

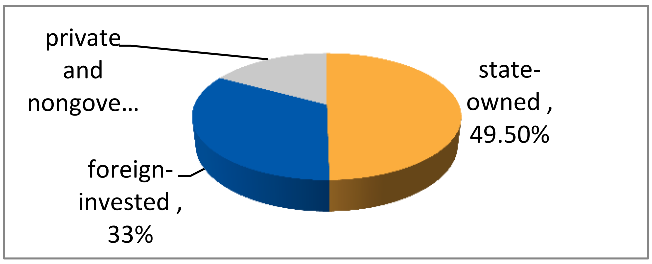

Figure 1-7: Composition of the sales volume of nationwide trans-city EDS

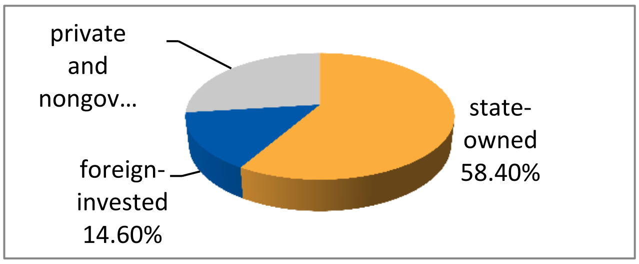

Figure 1-8: Composition of the business volume of nationwide trans-city EDS

(Source: based on the first statistic survey on EDS (2007))

State-owned enterprises

Although the number of state-owned enterprises, compared with that of private and nongovernmental enterprises, is quite small, most of them are well funded by the government and supported by different administrative departments, such as the Ministry of Railways and Civil Aviation Administration of China. Based on the unique resources and infrastructures they control, they have corresponding overwhelming advantages in trunk transportation. They cooperate with local EDS providers to offer nationwide door-to-door EDS. Enterprises such as SINOTRANS, CRE and CAE are of this kind.

[26] See the first statistic survey on EDS (2007).

Nongovernmental and private enterprises

There are more than 6000 nongovernmental and private EDS enterprises in China, let alone those service providers who have not registered with the government[27]. More than 90% of them constrain their service in the local market, i.e. within an economical region such as delta region of Yangtz River, within a city or even within an industrial park. Nongovernmental and private enterprises that offer nationwide trans-city EDS mostly start their business from a regional market and expand their service network nationwide with the advantages of high reputation in the regional market and high operational flexibility. Enterprises that have done quite a good job and are thus nationally well known to the public are very limited, taking SF Express and SHENTONG Express (STO) as examples.

International express magnates

After China's entry into the WTO, international express magnates, such as UPS, DHL, FedEx and TNT, consolidated and expanded their EDS in China through acquisition of Chinese homegrown enterprises or construction of the service networks by themselves[28]. Taking advantages of their advanced IT system, fully-fledged management and sufficient capital, they focus on the high-end market, by not only offering reliable services but also charging low prices, thus capturing more and more market share of the nationwide EDS.

[27] See the first statistic survey on EDS (2007).

[28] We collect the information from home page of corresponding enterprises and compile as follows. All the information is updated till Nov. 2012.

DHL announced on Mai 10, 2004 that it began to offer nationwide trans-city EDS officially in China. In order to expand its service coverage, in 2009 it purchased a native private company- Quanyi Express who had a widespread and profound service network. Service network of DHL in China now covers 318 cities, 123 of which are completely invested by DHL.

In 2006 FedEx reinforced its ground network in China by purchasing the left 50% share in Datian Logistic, whose service network covered 89 cities and regions in China. FedEx officially entered into Chinese EDS market in Mai. 28, 2007 and began to offer Next Midday EDS among 19 cities in China by cooperating with a nongovernmental airline called Aokai Airline. FedEx locates its air hub in Xiaoshan Airport. Now it has service points in more than 220 cities in China and plans to cover another 100 cities in next 4 to 5 years.

In order to develop EDS and logistic service in China rapidly, TNT purchased Huanyu, a leading LTL carrier in China in 2006. It has consequently a transportation network covering more than 1200 cities with more than 2000 service centers.

UPS began to offer internal parcel delivery service in China from 2005 through cooperation with Yangtze River Express.

Enterprise	Profile of network	Number of covered cities[29]
Company A	more than 45000 sales agencies, nearly 100 thousand employees	31 provinces and nearly 2000 cities
DHL	nearly 200 offices, about 7000 employees	no information
UPS	33 service centers	about 120
FedEx[30]	118 affiliates	224, another 100 are planned
TNT	more than1600 operation sites and 21000 employees, only ground transportation	more than 600
CRE	26 daughter companies, 2030 agencies	1317
CAR	33 affiliates, 3236 nationwide flight routes based on 140 airports, affiliated ground transportation networks	about 300
SF Express	more than 2200 points of sales	in 31 provinces, nearly 250 large and medium cities and 1300 country-level cities
STO	600 direct franchisees and more than 2000 indirect franchisees, about 4000 points of sales, 50 distribution centers, 40 thousand employees	nearly all prefectural-level cities
YT Express	56 distribution centers, more than 60 thousand employees	more than 1300
ZJS Express	32 affiliates, 3000 points of sales, 1000 agencies	more than 2000 cities and regions

Table 1-1: Network profile and service coverage of major nationwide trans-city EDS providers in China

(Source: mainly based on home pages of corresponding enterprises, update in Nov, 2012)

[29] We collect these data from home pages of corresponding enterprises and other official publications, which have different definitions of "city". In China, we have municipality directly under the central government, prefectural-level city and county-level cities. Since not all publications provide clear definition on "city" they cite, we only list the number they mention.

[30] Information is updated till 2008.

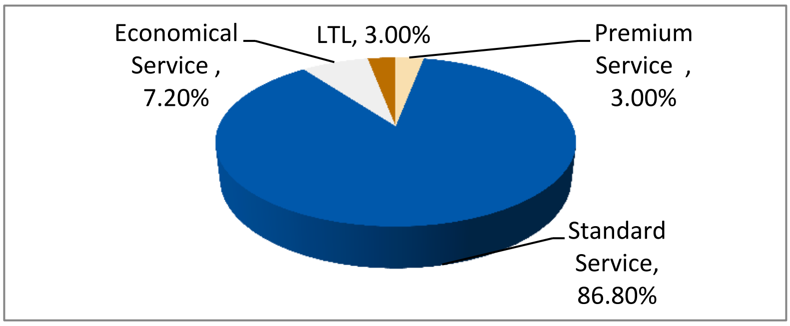

Figure 1-9: Express business composition of Company A

(Source: based on annual report of Company A in 2008)

Fig.1-9 shows the express business composition of Company A in 2008. As we can identify, the primary business of Company A is a standard express service. Specifically, it includes "next day EDS between major cities" and "2-3 day EDS across most parts of China". Compared to Company A, its market competitors, such as SF and DHL, take "next morning EDS between major cities" (premium service defined in Company A's current service system) as their primary business. In other words, the primary business of Company A is targeted to the mid-range and low-end market.

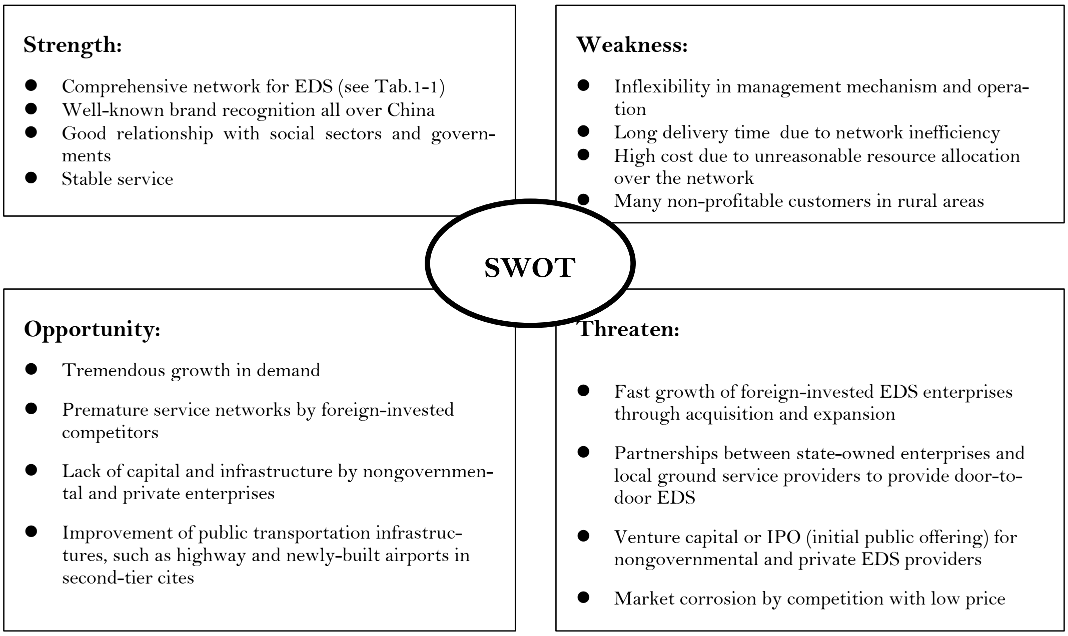

Figure 1-10: SWOT analysis of nationwide trans-city EDS of Company A

A profound SWOT analysis is conducted to help Company A position its target service in the nationwide trans-city EDS market (see Fig.1-10). Michael Porter's theory "Competitive Advantage" illustrates that in order to create a defensible status and outperform competitors in a given industry, firms should orient towards the specific generic strategies- overall cost leadership, differentiation and focus[31]. In logistic industry, it is

[31] See Porte (1998), p11.

practicable to carry out a hybrid strategy of cost leadership and focus, since the effect of EOS can be easily achieved[32].

Accordingly, Company A decides to focus on the high-end service in the primary market. To be more specific, Company A offers EDS of different service qualities simultaneously and focuses on high-end service in primary market. By focusing on the high-end service while offering standard service, Company A could serve its customers with lower cost, since the standard service can share the high fixed cost of the network and achieve economy of scale (EOS) and economy of scope (EOP).

1.3.3. Service positioning

New service system for nationwide trans-city EDS is defined according to the new market strategy (see Tab.1-2). Company A will focus on the key services, i.e. overnight EDS, while offering the standard EDS simultaneously. The network planning in this dissertation only involves the key services.

Service Type		Delivery time	Weight	Price level	Importance
Overnight EDS	Next Morning	collected before cut-off time, delivered before 12:00 next business day	<5kg	premium	key service
	Next Day	collected before cut-off time, delivered before 18:00 next business day	<30 kg	medium	key service
Standard EDS		collected before cut-off time, delivered within 2 or 3 business days	<30 kg	economical	normal service

Table 1-2: New service system of Company A

1.4. Outline of the dissertation

This dissertation is oriented towards the strategic planning of large-scale, multi-modal and time-definite networks for overnight trans-city EDS based on existing networks. The outline of the dissertation is illustrated in Fig.1-11.

Chapter 1 is the preparation for our research. In Sec.1.1 we briefly introduce EDS and its development. In Sec.1.2, we introduce the project background and clarify our research boundary in the overall project framework. In Sec.3 we define the new service system with marketing instruments based on the market strategy of the company. We also point out the target services for the network planning studied in this dissertation.

Chapter 2 demonstrates the current research gaps and research focuses of this dissertation through literature review on network planning for EDS. In Sec.2.1 we first illustrate planning levels and common features of EDS networks. In Sec.2.2, we make literature review on HLPs in perspective of the advance, taxonomy and conventional assumptions. By indicating current research gaps, we point out the category of our models and research focuses of our study in Sec.2.3.

[32] See Pfohl (2004), p.90.

Chapter 3 is devoted to the mathematical formulation of the models. In Sec.3.1 we describe in detail the network structure, parcel paths and service policies. In Sec.3.2 we formulate the basic model with linear air cost rate. In order to model the air cost more correctly, we study flow-dependent cost function in Sec.3.3. In Sec. 3.4 we propose two extension models with a cost select function for air backbone network.

Chapter 4 proposes the framework of the solution process, algorithms for different decisions and five customized improvement techniques. After literature review on solutions of relevant HLPs in Sec.4.1.1, we decide to adopt meta-heuristics, particularly hybrid GAs, to solve our models. In Sec. 4.1.3 we divide our original problem into three hierarchical sub-problems and propose a framework for the overall solution process. From Sec.4.1.4 to Sec.4.1.6 we discuss algorithms for hub location, demand allocation and air service selection decisions, respectively. In Sec.4.2 we propose five improvement techniques to different processes of GAs in order to improve the performance of the algorithms.

In Chapter 5 computational tests with public data sets are carried out with the basic model to evaluate the performance of the proposed hybrid GAs and the first four improvement techniques. In Sec.5.1, we test the performance of the overall algorithms under small-scale instances with the CAB data set by comparing its solutions with the corresponding optimal solutions generated by CPLEX. Sec.5.2 involves computational tests under large-scale instances with the AP data set. We modify the AP data set in Sec.5.2.1, set parameters for GAs with preliminary computational tests in Sec.5.2.2 and test the performance of the first four proposed improvement techniques in Sec.5.2.3.

Chapter 6 is dedicated to empirical study under real-life instances. Sec.6.1 specifies the preparation of input data for the models in the project. In Sec.6.2 we provide the solutions with extension model 1 (*Ext.1*) and extension model 2 (*Ext.2*) under the basic instance of the project data set and make some analysis and comparisons. We also conduct computational tests to see if the algorithms for *Ext.2* can be further improved with Improvement technique 5. In Sec.6.3 Scenario planning is conducted to help decision-makers identify critical factors, capture uncertainty, weigh between costs and corresponding decision risks in order to make robust decisions.

Chapter 7 summarizes the research and contribution of this dissertation. Equally important are the limitations and corresponding recommendations for future research.

Figure 1-11: Outline of the dissertation

2. Literature review on network planning for EDS

This chapter is devoted to literature review on network planning for EDS. In Sec.2.1 we illustrate the planning levels and common features of EDS networks. In Sec.2.2 we make literature review on HLPs in perspective of the advance, taxonomy and conventional assumptions. By indicating current research gaps, we point out the category of our models and research focuses of our study in Sec.2.3.

2.1. Network planning for EDS

2.1.1. Planning levels of EDS networks

The task for an EDS provider is to deliver mails and parcels from multiple origins to multiple destinations within a specified time period, which is called time-definite delivery service. On one side, customers in China constantly claim EDS of higher quality. In particular, they put great emphasis on delivery time and service reliability. On the other side, EDS providers have to face fierce market competition, increasing costs of fuel, toll and labor and an enhanced desire to protect the environment. To meet these challenges, they need an efficient service network.

For EDS providers that cover service region as large as China, multi-modal transportation networks are indispensable. They operate large-scale transportation systems that are composed of aircraft, vehicles, consolidation centers, sorting equipment and personnel to delivery mails and parcels between shippers and consignees. Thus, the network planning for EDS involves decisions about facility location and capacity, vehicle capacity, service quality (such as service frequency and delivery time), transport mode, vehicle routing and scheduling. In some cases, it also includes decisions about the reposition of empty containers and vehicles[33]. Traditionally, the planning of EDS network is carried out at three different levels in perspective of time horizon, i.e. strategic, tactical and operational planning. The following classification relates mostly to Teodor Gabriel Crainic[34].

- Strategic planning: Strategic or long-term planning looks several (sometimes dozens) years into the future. The planning is not constrained by resources at hand. It involves decisions on the physical structure with regards to resources, locations and infrastructure. The management decides e. g. where terminals and hubs shall be built, the volume of personnel to be employed, and how many vehicles of which type shall be bought. The strategic planning level sometimes also deals with the definition of customer service types and tariff policies.[35]
- Tactical planning: Tactical or medium-term planning deals with problems spanning from several weeks to several months with the aim of the design of the transportation network for the carriers. Given service policies, facility capacity and a finite number of vehicles and aircraft, tactical planning involves a set of interrelated decisions to optimally allocate and utilize resources to achieve the economic and customer service goals of the company[36]. Main decisions made at the tactical level concern the following issues: service

[33] See e.g. Jansen et al (2004), pp.41-53.

[34] See Crainic / Laporte (1997), p.409; Cranic (2000), pp.272-288; Cranic (2003), pp.451-516;

[35] See Wieberneit (2008), p.80.

[36] See Crainic (2000), p.272.

selection, traffic distribution, terminal policies and general empty balancing[37]. It is to be mentioned that researches under the name Service Network Design Problem (SNDP) [38]or Express Shipment Service Network Design Problem (ESSNDP)[39] actually deal with tactical and operational planning problems.

- Operational planning: Operational or short-term planning involves day-to-day decisions in a highly dynamic environment where the time factor plays an important role. It is based on the output of the tactical planning. The planners (usually local management, dispatchers) are confronted with a dynamic environment, where the orders may arrive dynamically or the time windows for pickup and delivery alter from customer to customer. It includes the implementation and adjustment of schedules for vehicles, crews and maintenance activities, and the control of the shipment.

Besides the three above-mentioned planning levels, there is a newly-proposed planning level called contingent planning[40]. While the traditional planning levels are oriented towards day-to-day situation, contingency planning aims at events of small probability. It prepares the system quick reaction to and recovery from accidents or even disasters, such as traffic accident, sudden demand volume change, weather disruption, breakdown of equipment, etc. It has become an increasingly hot topic in research circles to handle uncertainty or even disasters. With regard to this, planning levels of EDS networks can be summarized in Fig.2-1.

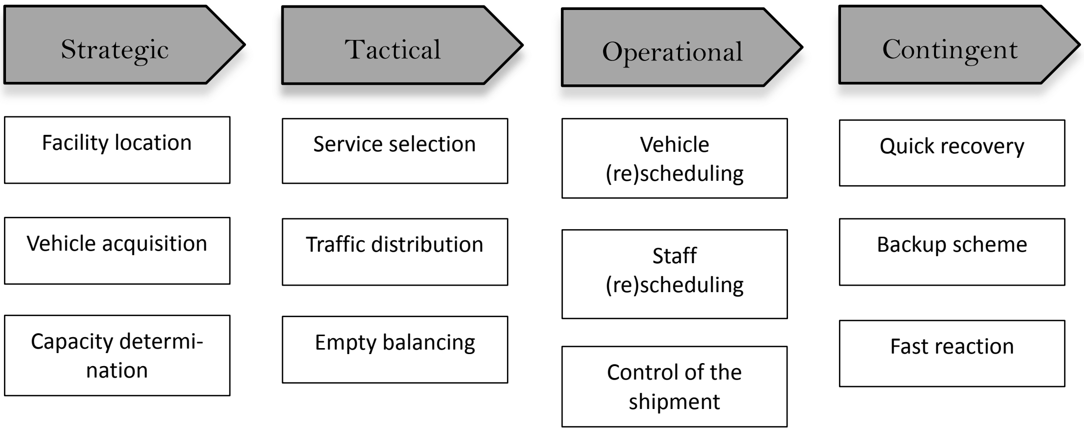

Figure 2-1: Planning levels of EDS networks

---

[37] See Crainic (2000), p.275.

Service selection: The routes on which services will be offered and the frequency and characteristics of each service.

Traffic distribution: The itineraries (routes) and vehicles used to move the traffic of each demand: services used, vehicles used, terminals passed through, operations performed in these terminals. But actually routing planning can belong to either tactical or operational planning level, depending on the planning horizon. For example, routing planning can take into account the interactions between different transportation models over a medium term planning horizon. See e.g. Crainic et al (1984), pp.165-184; Roy et al (1992), pp.31-44; But routing planning can also be for short planning horizon, e.g. five working days. See Mourgaya/ Vanderbeck (2007), pp.1028-1041.

Terminal policies: General rules that specify for each terminal the consolidation activities to perform.

General empty balancing: It indicate how to reposition empty vehicles to meet the forecast needs of the next planning period.

[38] See Crainic (2000), p.272; Wieberneit N. made a review on this problem. See Wieberneit (2008), pp.78-112.

[39] See Barnhart et al. (2002), p.239; Armacost et al. (2002), p.2; Kim (1997), p.685; Barnhart (1997), p.391; Grünert / Sebastian (2000), p.290; Büdenbender et al (2000), p. 364.

[40] See Barnhart et al. (2002), p. 244.

2.1.2. Common features of strategic network planning for EDS

In order to fulfill delivery tasks over a large area in a tight time window with minimum cost, EDS providers need time and cost efficient transportation networks. We list some common features of network planning for EDS, mainly at strategic level, by reviewing case studies on this or similar topics[41].

- Network structure

Network for multi-commodities[42] or many-to-many transportation service is commonly with a hybrid hub-and-spoke (H/S) configuration[43]. A typical H/S network consists of several tributary networks that connect demand nodes to hubs and a backbone network that connects the hubs. Depending on application domains, tributary networks are also called "local", "feeder" or "access" networks. Backbone networks may sometimes be referred as "long haul", "global" or "hub" networks. Hubs are named as "switches", "gateways", "control points" or "access points".

The impetus of applying H/S structure to multi-commodities networks comes from the considerations of cost, efficiency and operational flexibility. It is well-acknowledged that the H/S configuration is suitable for networks, in which it is expensive or impractical to establish a direct link between each origin-destination (O-D) pair. Traffic from different origins is consolidated at hubs and transported in bulk between hubs, since increased volume of inter-hub traffic brings economies of scale (EOS). Barges, vessels or larger trucks are used on inter-hub links for cost efficiency, while aircraft are for time efficiency. Moreover, when the H/S system is a nonrestrictive one[44], it has more flexibility for certain special requirement. Fig.2-2 shows a typical multi-allocation nonrestrictive H/S network in practice with the background of the US. Solid lines represent one-hub-stop and two-hub-stop service routes, and the dashed lines represent the nonstop service routes.

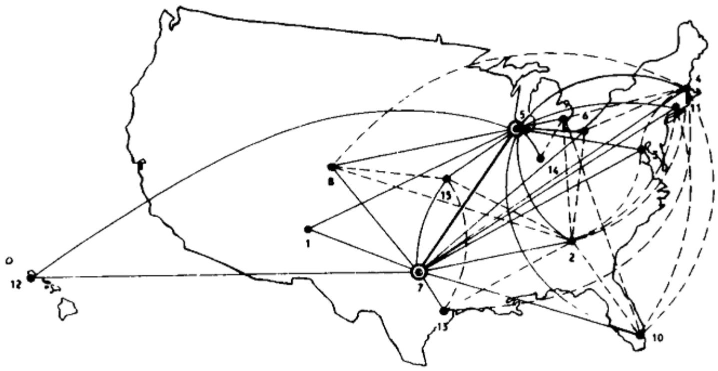

Figure 2-2: Multi-allocation nonrestrictive H/S transportation network in practice
(source: based on Aykin (1995), p.204)

- Objective functions & constraints

[41] We also go through case studies on network planning for postal service and less-than-truck service. We pay closer attention to case studies that are of real-life applications, are multimodal or cover large geographical area.

[42] All shipments from the same origin to the same destination with the same service level can be defined as a commodity.

[43] Hybrid H/S network refers to network with stopover in spoke network or with direct link between two non-hub nodes.

[44] In nonrestrictive H/S network, there are direct links between any non-hub nodes.

Providers of premium EDS always pay more attention to service quality, most importantly, delivery time. The latest arrival hub location problems are studies accordingly, taking the minimization of the maximum delivery time as the objective and the number of hubs as constraints.[45]

However, nearly every EDS provider should be responsible for its own profit and loss. Cost minimization is always regarded as the objective function in the strategic EDS network planning under the constraints of service quality. Since service providers can capture their market share by offering similar services as their competitors while charging lower prices.

- Problem size & solution methods[46]

Problems in real-life always mean large instance scale. Aggregation and clustering of demand nodes are two effective methods to reduce the instance scale. For example, in case of network restructuring for Swiss parcel delivery service, 3,700 Swiss postcode regions are aggregated to about 200 demand points as input of the model[47].

Even though, problems are still so large that heuristics or meta-heuristics, which can handle large-scale problems, are applied, although optimal solutions are not guaranteed. For example, an iterative hubbing –routing heuristic is applied to a network in Turkey with 81demand nodes.[48] In the case of configuring a H/S network for a less-than-truck (LTL) trucking company in Brazil, genetic algorithms (GAs) is applied on HLPs with 46 demand nodes and 46 potential hubs.[49]

2.2. Literature review on HLPs

The strategic network planning for EDS in this dissertation will simultaneously determines hub location, demand allocation and air service selection. In this section we make literature review on HLPs that involve the first two decisions, i.e. hub location decisions and demand allocation decisions. The last type of decisions, i.e. air service selection decisions, actually aircraft fleet ownership decisions, will be discussed in Chapter3.

2.2.1. Participants in H/S networks

- Carriers

The primary motivation for carriers to adopt H/S network structure is to seek cost efficiency from EOS[50]. For one thing, by consolidating flows through hubs inter-hub transportation cost rate by larger vehicles is lower than the transportation cost rate of moving commodities directly from origin to destination by smaller vehi-

[45] See e.g. Kara / Tansel (2001), pp.1408–1420.

[46] For details about solution methods for HLPs, please see Chapter 4.1.1.

[47] See Bruns et al. (2000), pp.285–302.

[48] See Cetiner et al (2010), pp.109-124.

[49] See Cunha/ Silva (2007), pp.747-758.

[50] See Pfohl (2004), p.127.

cles[51]. For another, by bundling flows H/S networks need much less links to serve all O-D pairs than fully interconnected networks [52]. H/S networks with feeder routes and stopovers need even less vehicles than pure H/S networks[53]. However, this cost efficiency is based on the premise that the cost saving from bundling and consolidation can compensate the cost increase resulting from detour and transshipment[54].

With regard to management, hubs, as transshipment nodes for all shipments, become the core of whole system. Maintaining smooth flow at hubs is a tough but critical task for service providers. The reliability of the H/S system is more sensitive than fully interconnected network, since the paralysis of hubs will be a disaster to the whole system rather than only breakdown of a part of O-D pairs.

- Network users

H/S networks have both positive and negative impacts on their users. On the one hand, the cost saving from H/S networks for the carriers is partly at the cost of users, although it must not be a zero-sum game. For air passengers, it means longer travel time, inconvenient stopover at a third airport or even higher probability of luggage delay. For air freight, it means longer delivery time and more risk from transshipment. Therefore, users must be compensated with low fares. Otherwise they will turn to other direct service providers.

On the other hand, H/S network adopter usually benefits users with high service frequency, which obviously increases the network access[55]. This strategy can make up part of its deficiency and help the service provider to seize market share[56].

2.2.2. Advance of HLPs

The research topic HLP originated from aviation industry. After the Air Cargo Deregulation Act was issued in 1978 by Civil Aeronautics Board (CAB) in the U.S., most airlines began to transform their air networks to H/S structure, which invoked academic research in this area[57].

First, studies put more emphasis on market strategy, profitability of airlines and passenger welfare (such as frequency of flights and fares)[58]. In most cases air hub location problems came under the research area of eco-

---

[51] A similar research topic of HLPs is facility location problems (FLPs). The fundamental difference between these two topics is link between facilities. In FLPs, service is offered at or from the facilities. So the network planning problems determine where to locate the facilities and how to connect demand nodes to their "home" facilities. However, demand in HLPs is specified as flow between origin and destination (e.g. flow of passengers, information or commodities). The facilities (here hubs) serve as consolidation nodes along the flows that connect pairs of O-D. Although these two problems are quite similar in other aspects, there is no EOS on hub links in FLPs.

[52] Daskin illustrated this mechanism by a sample network with six nodes. See Daskin (1995), p.3. To be more general, if we have N nodes and if each node can be either origin or a destination, we need $N(N-1)$ direct links in a network to compose a fully interconnected network. If we designate one of the nodes as hub and connect it to the rest nodes, we need only $2(N-1)$ links to serve all O-D pairs. The saving becomes larger as N increases.

[53] See Lin/Chen (2004), pp.271-283; Kuby /Gray (1993), pp. 1-12.

[54] See Domschke/ Krispin (1999), pp.279-304.

[55] See Butler/Huston (1990), pp. 3-16.

[56] See e.g. Borenstein (1989), pp. 344-365; Borenstein (1991), pp.1237–1266; Berry (1990), pp.394–399.

[57] See Jaillet et al (1996), p.195.

[58] See Morrison/ Winston (1986).

nomics. Air hubs were selected with regression models or other econometric models by considering a number of economic factors, such as population, per capita income and Gross Domestic Product (GDP)[59].

Later, the topic HLP was introduced by a number of pioneering researchers into the domain of management science (MS) and operations research (OR). More attention was paid to network structure (i.e. hub number and demand location), total costs and constraints on service quality and capacity. We confine the following literature review on HLPs within the domain of MS and OR.

O'Kelly[60] presented in 1987 the first recognized mathematical formulation for HLP by studying a passenger airline network. This pioneering work attracted the attention of researchers from a wide variety of fields. Since then HLP has become an active topic in MS. Campbell[61] played a major role in completing modeling on HLPs. His papers are among the most important studies on hub modeling. Some authors also made great contributions to improving this topic, including Aykin[62] and Klincewicz[63]. In an effort to organize the growing number of papers on HLPs, O'Kelly & Miller[64], Daskin[65], Skorin-Kapov & Skorin-Kapov[66], ReVelle[67] , Klincewicz[68] and Bryan & O'Kelly [69] made literature reviews on this topic.

Early researches were based on rather strict assumptions on the network. The problem scale was quite small, probably owing to the limitation of computer technology at that time. Recognizing that the complexity of the HLPs prevented many important characteristics of real-life H/S networks from being modeled, some researchers simplified the problem by holding the hub locations fixed, so that they could pay attention to incorporating more realistic characteristics of hub networks into models. Such extensions included the use of direct links between non-hub nodes[70], mini or regional hubs for EDS systems[71], network planning with profit maximization as objective[72], and congestion problems at hubs[73].

However, one must keep in mind that hub location decision is always the key issue for strategic network planning. As computer technology improved and knowledge on HLPs grew, hub location and other decisions began to be made simultaneously. Several researchers developed models that incorporated additional important characteristics within the HLP framework. For example, models with hub fixed costs, capacity constraints to

---

[59]See Bauer (1987), pp.13-19.

[60]See O'Kelly (1987), pp. 393–404.

[61] See Campell (1994), pp.31-49.

[62] See Aykin (1994), pp. 501-523; Aykin (1995), pp.201-221.

[63] See Klincewicz (1991), pp. 25-37; Klincewicz (1992), pp. 283-302.

[64] See O'Kelly et al (1994), pp. 31-40.

[65] See Daskin (1995), Chapter8.7.

[66] See Skorin-Kapov/ Skorin-Kapov (1995), pp.183-192.

[67] See ReVelle (1997), pp.3-13.

[68] See Klincewicz.(1998), pp.307-335.

[69].See Bryan/ O'Kelly (1999), pp. 275-295.

[70] See O'Kelly/ Miller (1994), pp. 31-41; O'Kelly (1998a), pp. 171-186; Jeng (1987); Flynn/ Ratik (1988), pp. 139-147; Kuby/ Gray (1993), pp. 1-12.

[71] See Hall (1989), pp.139-149; O'Kelly/ Lao (1991), pp. 283-297; O'Kelly (1998b), pp.77-99.

[72] See Daskin/Panayotopoulos (1989), pp. 91-99; Dobson/ Lederer (1993), pp. 281-297.

[73] See Grove/ O'Kelly (1986), pp. 103-119.

reduce congestion at hubs, threshold constraints to prevent links from being underutilized and modified cost function for inter-hub links[74]. Moreover, the objective of minimizing network cost may not be appropriate for all applications. Research efforts were also directed towards designing networks with various objectives. For instances, the *p*-hub center problem was proposed to minimize the maximum service time. Maximal hub covering problem was to cover as many demands as possible with a predefined number of hubs.

Some recent reviews on HLPs summarize in detail the research status quo[75]. The statistics by Hekmatfar M. and Pishvaee M.[76] show that the number of published papers on HLPs has increased significantly since 1985. We extend their work by updating the data till recently (see Fig.2-3). We collect the data with the search engine "web of knowledge"[77]. We search papers with the key words "hub location", "hub and spoke" and "network planning" and include the papers that take HLPs as the main topic by reading through all the abstracts. We must note that the only one database has limited coverage. However, the purpose of this work is to distinguish the advance trend of research on HLPs. Our results for the years 1985-2006 are quite similar to those by Hekmatfar M. and Pishvaee M., with small discrepancies in several years.

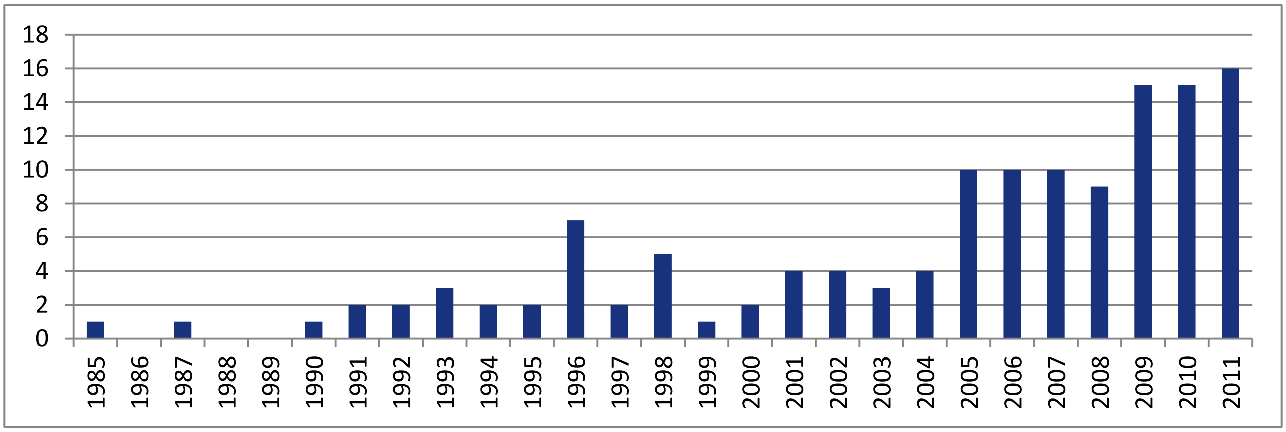

Figure 2-3: Annual number of papers on HLPs (compiled by author)

Research focus of HLPs was on modeling in early years, on improvement of models in the following years and on solution techniques in recent years[78]. Moreover, three obvious trends have appeared in recent works, i.e. more researches on multi-level network planning problems, on hierarchical network planning problems and more considerations on time constraints.

Multi-level network planning problems

[74] See e.g. Chou (1990), pp.243-258; O'Kelly (1992), pp. 293-306; Campbell (1993), pp.473-482; Aykin (1994), pp.501-523; Campbell (1994b), pp.31-49; Akyin (1995), pp.201-221; Jaillet et al. (1996), pp.195-212; O'Kelly (1998a), pp. 171-186; O'Kelly (1998b), pp.77-99; O'Kelly/ Bryan (1998), pp. 605-616.

[75] See Hekmatfar /Pishvaee (2009), p.247; Thomadsen / Larsen (2007), pp.2520-2531; Alumur / Kara (2008), pp.1-21; ReVelle /Eiselt (2005), pp.1-15; ReVelle et al. (2008), pp.817-814.

[76] See Hekmatfar /Pishvaee (2009), p.247.

[77] URL: http://wokinfo.com/..

[78] See Hekmatfar /Pishvaee (2009), p.247.

Multi-level network planning problems[79] here refers to the expansion of conventional HLPs from strategic planning to compound strategic-tactical planning by incorporating routing problems into location problems. It had been found that the conventional H/S network structure, i.e. fully interconnected/ star shaped structure, is effective in reducing transportation cost. But additional cost saving effects could be achieved if a tributary trip starts from a hub and covers many customers to make a tour and if a backbone trip covers many hubs. It would definitely reduce the number of vehicles operating in the system. With regards to this, locating hubs and generating multi-stop routes for tributary and backbone networks are combined together as hub location-routing problems (HLRP)[80].

In addition to identifying the location of hubs and the allocation of customers to the hubs in conventional HLPs, HLRPs must determine the allocation of customers to the routes, the order of visiting customers in tributary routes and the order of visiting hubs in backbone routes. Hence, HLRPs contain both location problems and vehicle-routing problems (VRRs)[81]. HLRP is essentially a location problem with the distinguishing property of paying special attention to underlying issues of VRP. In order to achieve the optimization of the location problem (master problem), VRP (sub-problem) must be simultaneously considered[82]. The networks for conventional HLPs and HLRPs are compared in Fig.2-4 and 2-5 respectively.

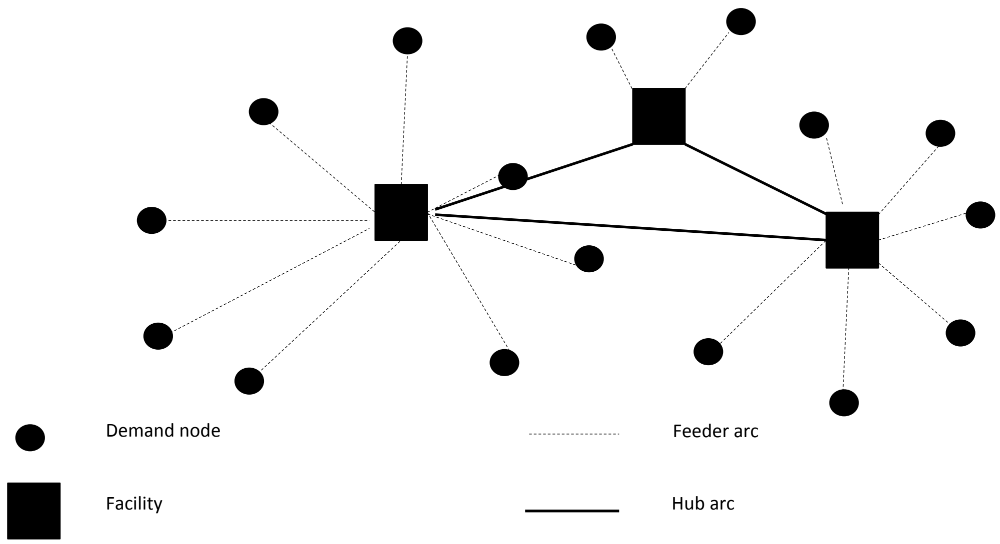

Figure 2-4: Network for conventional HLPs

[79] The literal synonyms "level", "layer", "stage" and "hierarchy" here imply different meanings. The word "level" is used to distinguish planning horizon in this dissertation, namely strategic, tactical, operational and contingent planning as defined in Sec.2.1.1, while the words "layer", "stage" and "hierarchy" in the next bold tip are synonymously applied to describe network structure or architecture.

[80] We follow the name proposed by Cetiner et al. (2010), pp. 110. However, the HLRP defined by the author includes only multi-stop routes in tributary network but not in backbone network.

[81] The VRP is concerned with the determination of the optimal routes used by a fleet of vehicles to serve a set of customers. See Min et al (1998), pp.1-15; Lin/ Kwok (2006), pp.1833-1849; Nagy/ Salhi (2007), pp. 649.

[82] It was formerly proposed to LRP. But it is also applicable to HLRP. See Nagy/Salhi (2007), p.650.

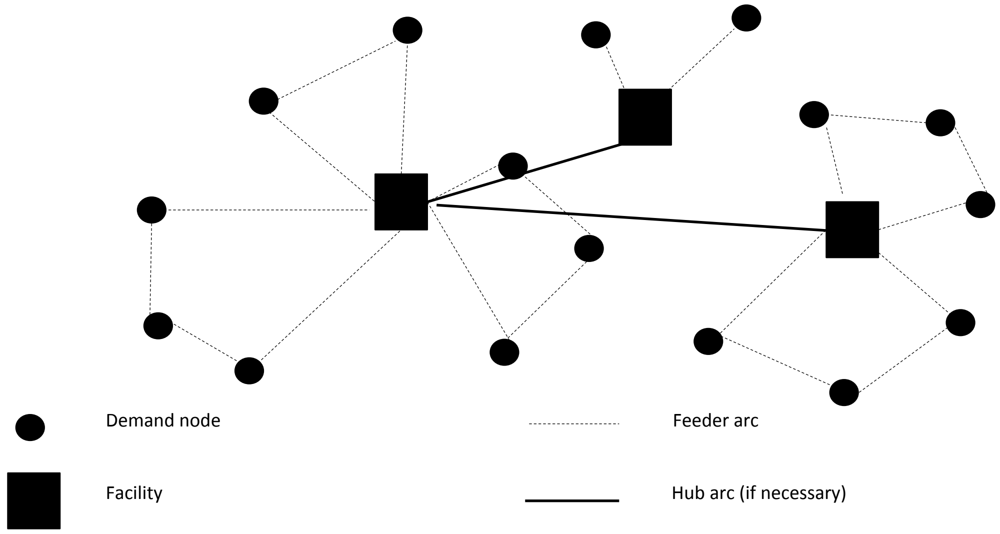

Figure 2-5: Network for hub location-routing problems (HLRPs)

Actually, similar compound problems appeared first in FLPs as location-routing problems (LRPs) [83] .LRPs pay special attention to the underlying VRPs within FLPs. In addition to identifying the number and the location of facilities, LRPs also determine the allocation of customers to the facilities, the allocation of customers to the routes, and the order to visit the customers in routes. The popularity of research on LRPs almost parallels the advent of an integrated logistics concept[84]. Compared with LRPs, routing problems in HLRPs are not only involved in tributary networks but also in backbone networks. The relationship between conventional location problems and their expansions is displayed in Fig.2-6.

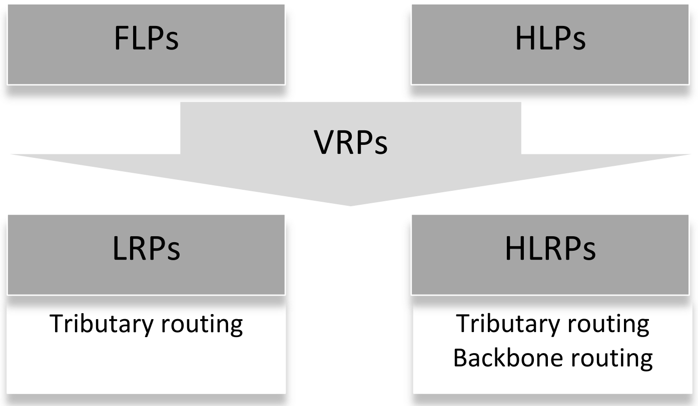

Figure 2-6: Extensions of conventional location problems

[83] See e.g. Or/ Pierskalla (1979), pp. 86–95; Jacobsen/ Madsen (1978), pp.378–387; Laporte / Norbert (1981), pp.224–226. For overview, please see Nagy/Salhi (2007), pp. 649-672.

[84] See Barreto et al. (2007), pp. 968-977; Sambola et al. (2005), pp.407-428.

To the best of our knowledge, routing possibilities within HLPs were first considered by Kuby and Gray[85]. Their formulation is to determine the least cost set of direct and stopover routes for the traffic from demand points to a hub. However, in their model the only hub is predetermined and there is no inter-hub link. Nagy and Salhi[86] are the first ones to study HLRP, simultaneously determining hub locations and multi-stop feeder routes. They proposed a hierarchical heuristic, in which hub locations are determined in a master problem and routings are determined with neighborhood search in a sub-problem. Bruns et al [87] investigated the restructuring of the Swiss postal services and the problem was composed of decisions on the number, capacity, and location of transshipment points. Although there was no routing problem, they estimated the routing costs for the location model. Wasner and Zäpfel[88] studied HLRP for parcel services. The model considers the possibility of direct transport between two non-hub points.

There are also HLRPs, in which hubs are connected with tours while customers are connected to hubs directly. For example, Maheshwari[89] developed a model for multi-zone truckload shipments. Direct shipments were assumed between a spoke and a hub, but loads could be sent through several hub-to-hub shipments. A construction heuristic with tabu search (TS) framework was developed to solve larger sized problems. The study by Labbe et al [90] is to determine a simple cycle through a subset of vertices of a graph involving two types of costs: routing costs associated with the cycle itself and costs of assigning vertices not on the cycle to visited vertices. The objective is to minimize the routing costs, subject to an upper bound on the total assignment costs. Gunnarsson et al [91] considered a combined terminal location and ship routing problem for pulp distribution in Scandinavia. The purpose is to satisfy customers' annual demand of pulp products while minimizing the distribution costs.

More complicated HLRPs include those with tours occurring at both hub and customer levels. Melechovsky et al[92] dealt with a location-routing problem with non-linear cost functions. The proposed heuristics algorithm is a combination of a p-median approach to find an initial feasible solution and a meta-heuristic to improve the solution. It is a hybrid meta-heuristic merging Variable Neighborhood Search (VNS) and TS principles.

Hierarchical network planning problems

Hierarchical network in this dissertation can be also named as multi-layer or multi-stage network. One layer means one type of consolidation facilities in the network, either primary layer, secondary or even third layer.

Conventional studies on HLPs are primarily concerned with one-layer models that include hubs and demand nodes. However, recent studies are more likely to extend the network span by considering two layers or even three layers in the network. That is to say, the network includes more than one kind of consolidation facilities,

[85] See Kuby/ Gray (1993), 1–12.

[86] See Nagy /Salhi (1998), pp. 261–275.

[87] See Bruns et al (2000), pp. 285–302.

[88] See Wasner / Zäpfel (2004), pp. 403-419.

[89] See Maheshwari (2004), pp. 16-31.

[90] See Labbe et al. (2005), pp. 457–470.

[91] See Gunnarsson et al. (2006), pp.928-938.

[92] See Melechovsky et al. (2005), pp. 375-391.

for example hubs, depots, stations or centers. There are already a few HLPs dealing with two layers of facilities[93]. Those related to EDS are by Wasner & Zäpfel, Lin & Chen, and Yaman & Ben-Ayed[94].

Considerations on time constraints

With the development of HLPs, more considerations are included into models gradually. Service time becomes a main concern, which can be either the constraints in the model or the objective of the model[95]. This trend is triggered by requirement from customers and fierce market competition.

For example, hub set covering problems involve H/S networks, in which commodities are delivered within certain time window between all O-D pairs. The objective is to minimize costs for hubs to be opened.[96] The latest arrival hub network design problem, also called *p*-hub center problem[97], is the problem to determine the location of hubs, the allocation of non-hubs to hubs and the associated routes between non-hubs and hubs with multiple stopovers. The objective is to minimize the maximum delivery time.[98]

2.2.3. Taxonomy of HLPs

HLPs can be categorized according to different criteria[99], three of which are most commonly applied: objective function, allocation criterion and network architecture.

2.2.3.1 Objective function

As research branch of facility location problems (FLPs), almost all HLPs can find their counterparts in FLPs (see Tab.2-1). So the classification of FLPs according to objective function can be generally applied to HLPs[100].

We distinguish them into two categories, one with exogenous facility number and the other with endogenous facility number.

---

[93] The corresponding problem in FLP is e.g. the *pq*-median problem proposed by Serra and ReVelle, which seeks to locate hierarchical facilities at two levels so as to obtain a coherent structure. See Serra/ ReVelle (1993), pp. 299-312; Serra/ ReVelle (1994), pp.63-82; Alminyana et al (1998), pp.1-23.

[94] See Wasner / Zäpfel (2004), pp.403–419; Lin/ Chen (2004), pp.271-283; Yaman (2009), pp.643–658; Ben-Ayed (2010), pp.250–269; Ben-Ayed (2011), pp. 1-22.

[95] See Lin/ Chen (2008), pp. 986–1003; Chen et al (2008), pp. 493–515; Campbell (2009), pp.3107-3116.

[96] See Wagner (2007), p.932. Other researches on hub set covering problems, please refer to Campbell (1994), pp.387-405; Calik. et al (2009), pp.3088-3096; Alumar/ Kara (2009), pp.1349-1359. It is also discussed in Sec.2.2.4.

[97] See Kara/ Tansel (2003), pp. 59-64; Kratica/ Stanimirovic (2006), pp.425-437; Campbell, et al. (2007), pp.819-835; Ernst (2009), pp. 2230-2241; Campbell (2009), pp. 3107-3116; Calik et al (2009), pp. 3088-3096.

[98] See Yaman et al. (2007), pp. 906-919.

[99] Please refer to some recent reviews on HLPs. See Hale/Moberg (2003), pp.21-35; Alumur/Kara (2008), pp.1-21; ReVelle/Eiselt, (2005), pp.1–19; Campbell (1994b), pp.387-405.

[100] See ReVelle et al. (2008), pp. 817–848; Nagy/ Salhi (2007), pp.649-672.

		FLPs		HLPs
Exogenous P	*P*-median	Minimize the total transportation cost between demand nodes and facilities	*P*-hub median	Minimize the total transportation cost (including backbone network)
	P-center	Minimize the maximum distance or time between any demand node and its home facility	*P*-hub center	Minimize the maximum distance or time between any origin-destination (or origin-to-hub, hub-to-hub or hub-to-destination) pair
	Maximal covering	Maximize covered demand with *P* facilities	Maximal hub-covering	Maximize covered demand with *P* hubs
Endogenous P	Total covering	All demands are covered with least number of facilities	Hub set-covering	All demands are covered with least number of hubs
			Hub location with fixed cost	All demands are covered with minimum cost (both transportation and hub fixed cost).

Table 2-1: Classification of HLPs according to objective (with counterparts in FLPs)

However, HLPs are different from FLPs in three aspects, namely decision-making mechanism, allocation considerations and service region[101].

(1) Decision-making mechanism

The difference between some FLPs[102] (such as location of hospitals and supermarkets) and HLPs is that FLP involves a user attraction system, while HLP concerns a goods delivery system. This distinction was first mentioned by O'Kelly and Miller[103]. For delivery systems, the networks are planned from the perspective of carriers in terms of both location and routing problems. In such model it makes sense for the entire problem to be treated as a unified simple objective optimization task. The end users (or customers, i.e. shippers or consignees) have no interest in the path of the goods. Only the cost of the system may impact on the price charged for the service. In user attraction systems, in contrast, facilities are located by service providers, while the end users decide which facility to use. In other words, in some FLPs, location and allocation (routing) decisions are decentralized. The network planner has to make some reasonable guesses on how the public will make use of those facilities. So the inconvenience and consumer behavior cannot be ignored in some FLPs.

(2) Allocation considerations

The fundamental difference between HLPs and FLPs is that demand in FLPs is represented by a point or area, while demand in HLPs is represented by a pair of nodes. In other words, facilities must be interconnected in

[101] Some ideas were mentioned by O'Kelly. See O'Kelly (1998), pp.172-173.

[102] They are referred to certain kind of FLPs, in which customers are served at facilities. Facilities, such as fire station or newspaper distribution centers, do not belong to this kind.

[103] See O'Kelly/ Miller (1994), pp.31-40.

HLPs to serve all O-D pairs, while facilities are independent in FLPs. When the cost of traveling across the inter-hub link is free, HLPs reduce to FLPs[104].

For this reason, in some FLPs[105] demands are allocated with nearest-distance criterion, since the travel cost or travel time consists of a single component: the segment from demands to the facility. However, in HLPs with restrictive H/S structure[106] the travel cost or travel time of each demand consists of at least three components: (1) the travel cost from the origin to the hub, (2) the cost between hubs (if necessary) and (3) the travel cost from the hub to the destination. In this sense, the total travel cost must be considered for allocation decisions in HLPs.

(3) Service region

Nearest-distance allocation decisions in FLPs guarantee non-overlapping service regions. In turn, single allocation criterion is not appropriate for all H/S systems by considering time constraints. Service regions in HLPs are sometimes overlapping.

The comparison between HLPs and FLPs is summarized in Tab.2-2.

	HLPs	**FLPs**
Demand	O-D pair	Node or area
Decision-making mechanism	Delivery system with centralized decisions	User attraction system with decentralized decisions
Allocation considerations	Least travel cost	Nearest distance
Service region	Probably over-lapping	Non over-lapping

Table 2-2: Comparison between HLPs and FLPs

2.2.3.2 Allocation criterion: single or multi-allocation

How many hubs a customer/demand node can be served, one or more? The answer to the question implies whether the network adopts single allocation criterion or multi-allocation criterion. H/S networks with single allocation criterion require any demand node to be uniquely assigned to an exclusive hub, while H/S networks with multi-allocation criterion allow demand nodes to be connected to several or all hubs. Different allocation criteria, whether single or multiple, result from different planning consideration and organization structure, and result in different transportation costs, system flexibility and reliability.

Planning consideration

In most cases a hub network that adopts multi-allocation criterion takes users' benefit as one of the most important considerations, since the system may give up EOS on feeder and even backbone transportation to offer

[104] See O'Kelly (1987), pp. 393-404.

[105] FLPs here are referred to the kind of FLPs, in which customers are served at facilities.

[106] That is all demands must be consolidated at hubs.

service with less travel time[107]. An air passenger network is best represented as a multi-allocation network, in which each demand node is potentially linked to a variety of hub cities, and therefore a passenger has the option to pick up a more convenient hub, through which to make a transfer.

When cost efficiency of the system is under consideration, single-allocation models are a better choice for freight or communications networks, although sometimes multi-allocation models are also adopted by freight networks due to time or other constraints. The EOS on inter-hub links gives carriers the incentive to direct flows towards few hubs and make allocation decisions to minimize the transportation cost, while ignoring detours and inconvenience during the travel[108].

Organization structure

Organization structure should be in line with the structure of the transportation network. As a matter of fact, most LTL truck companies and EDS providers adopt single-allocation network, allowing for the non-overlapping supervision and clear responsibility between the hub regions [109].

Transportation cost

Under the multi-allocation criterion, the traffic originating from or ending at a certain demand node is divided into several parts that are transshipped through different hubs. Even the traffic with the same O-D can be divided and each part selects the path that minimizes its own travel cost. A single allocation model can be regarded as a special case of a multi-allocation model, when traffic can only go from the origin to its exclusive "home" hub and from "home" hub to the destination. Hence the variable transportation cost of multi-allocation models is at least as low as its corresponding single-allocation versions[110]. However, when fixed costs of spoke links are considered, this is not always the case. It will result in less spoke links in the network. And the higher the fixed cost is, the less multi-allocation happens[111].

System flexibility

Moreover, multi-allocation policy can serve as an effective cost-saving measure in face of rise on inter-hub discount rate (α). O'Kelly et al found that there is an increase in the number of multiple allocations as the inter-hub discount rate increases. The single-allocation model cannot employ this cost-saving strategy and may respond only by (1) changing hub location or (2) changing allocation decisions. Comparatively speaking, these two measures are not as efficient as the increase of multiple allocations that is only available to multi-allocation models. As a result, the discrepancy of total cost between the two models increases as the inter-hub discount rate increases[112]. In this respect, single-allocation systems are not as flexible as multi-allocation systems under of uncertainty of transportation cost rate.

[107] See O'Kelly (1998), p.174.

[108] See O'Kelly/ Morton (1998), p.171.

[109] See Bruns et al. (2000), p.290.

[110] See O'Kelly / Bryan (1996), p.126. Also see the result from Ernst/ Krishnamoorthy (1996), pp.139-154. and Ernst/ Krishnamoorthy (1998), pp. 100-112.

[111] See Jaillet et al. (1996), p. 210.

[112] See O'Kelly / Bryan (1996), pp.125-138.

System reliability

If it is permissible, or even required, that a demand node is connected to more than one hub, this provides alternate routes in case of a spoke link failure. When a hub breaks down in multi-allocation networks, the demands served by this hub can be partially or even totally served by other hubs. However, in single-allocation systems all the service for its subordinate nodes is interrupted. In this case, the negative impact on the multi-allocation system is much less than that on the single-allocation system.

The comparison of single allocation and multi-allocation criterion is summarized in Tab.2-3.

	Single allocation	**Multi-allocation**
Applicable service	Passenger network	Freight and communications network
Planning consideration	High efficiency in carrier's perspective	Convenience in travelers' perspective
Organization structure	Single supervision& clear responsibility	Over-lapping service region
Total variable transportation cost	Relatively high	Relatively low
System flexibility	Relatively low	Relatively high
System reliability	Relatively low	Relatively high

Table 2-3: Comparison between single-allocation and multi-allocation criterion

2.2.3.3 Configuration of the network

A H/S network is composed of one backbone network and several tributary networks, both of which can take different configurations summarized as follows.

Backbone network: fully interconnected /mesh/star/tree/ring

Tributary network: star/ tree/ path/ ring

In most cases backbone network is assumed to be fully interconnected. But if one or more of those backbone links are not presented, it would be a "mesh" network as illustrated in Fig.2-7. The tributary network associated with hub A is a so-called "star" network, in which every other node is directly linked to hub A. The tributary network associated with hub B is a "tree" network and the tributary network associated with hub C is a simple "path" network. Finally, the tributary network associated with D has a "ring" structure. In a ring structure, traffic can travel around the ring in two directions.

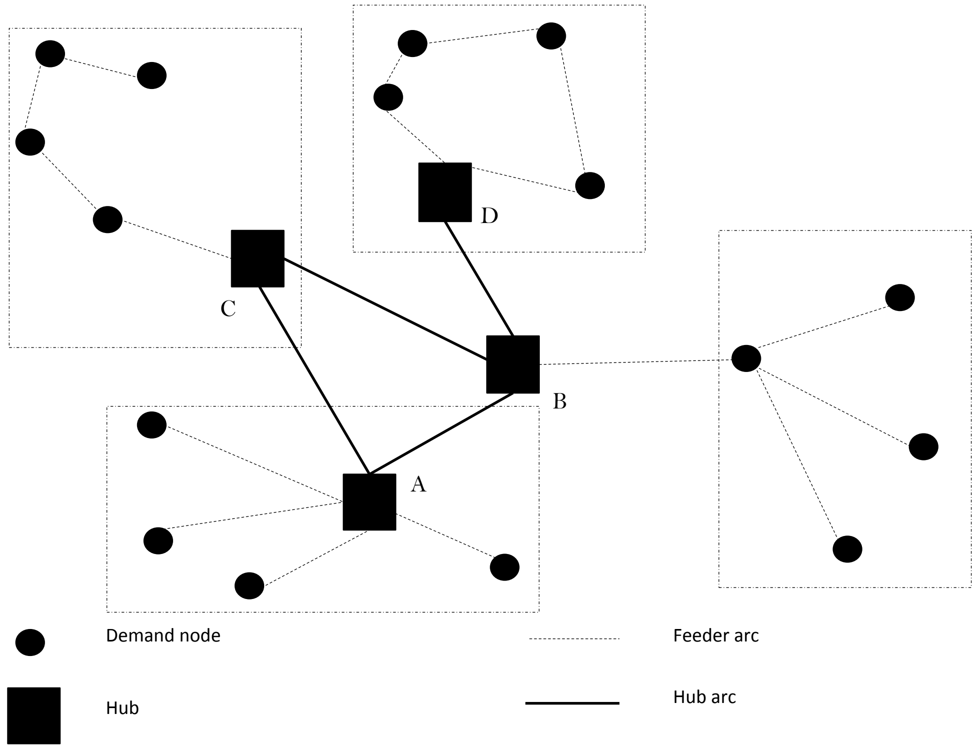

Figure 2-7: Illustration of backbone and tributary networks

So this can result in many different combinations for different modeling purposes. For example, star/star networks[113], tree/star networks[114], fully interconnected/star networks[115], mesh/ star networks[116], ring/star networks[117] and different backbone networks with tree tributary networks[118].

2.2.4. Review on classical HLPs

In this section we introduce in detail four classical HLPs. We take uncapacitated single-allocation ones as examples.

***P*-hub median problem**

[113] In star/star network, a number of demand nodes are connected directly to intermediate hubs; the intermediate hubs, in turn, are connected directly to a central hub. It is already a well-studied problem. See Minoux (1989), pp. 313-360; Balakrishnan et al. (1991), pp. 237-284; Lee (1993), pp. 471-482; Sridharan (1993), pp.305-312.

[114] It is a network, in which central hub can be connected to other intermediate hubs in a hierarchical fashion. See e.g. Pirkul/ Nagarajan (1992), pp.247-261; Chamberland et al. (1996), pp. 525-536.

[115] This network structure is a common one that has been well discussed. See e.g. Campbell (1994b), pp.1-19; Sohn /Park(1997), pp.617-622 .

[116] See e. g. Boorstyn/ Frank (1977), pp.29-47; Gavish (1992), pp. 149-172.

[117] See Current/ Schilling (1994), pp.114-126; Gendreauet al. (1997), pp. 568-576.

[118] Example for path/ tree, see Pirkul et al. (1991), pp. 175-182; Example for tree/ tree, see Balakrishnan et al. (1994), pp.567-581.

The objective of the p-hub median problem is to minimize the total transportation cost to serve all demand, given n demand nodes, demand flow between all O-D pairs and the number of hubs to be located (p).

Single allocation p-hub median problem was the first recognized mathematical formulation of HLP by O'Kelly[119]. Campbell[120] produced the first linear integer programming formulation for this problem by defining four-subscript variables. Since the LP relaxation of Campbell formulation results in highly fractional solutions, Skorin-Kapov et al.[121] proposed a new mixed integer formulation for the problem. Other formulations were also proposed by Ernst and Krishnamoorthy[122] and Ebery[123].

We introduce the p-hub median problem with the formulation by Skorin-Kapov et al.[124]. We adopt this four-subscript formulation method to formulate our own models in later chapter. Let w_{ij} be the flow between nodes i and j and c_{ij} be the transportation cost of a unit of flow between i and j. Define x_{ik} as 1 if node i is allocated to hub k, and 0 otherwise; x_{kk} takes on the value 1 if node k is a hub and it is 0 otherwise. Also define x_{ijkm} as flow from node i to j that is routed via hubs at locations k and m in that order. Parameter α is the factor for EOS; the unit cost of flow between hubs must be smaller than that between hubs and demand nodes since hubs concentrate flow, so $0 \le \alpha < 1$. The formulation is as follows.

$$Min \sum_i \sum_j \sum_k \sum_m w_{ij} x_{ijkm} (c_{ik} + c_{mj} + \alpha c_{km}) \tag{2-1}$$

$$\text{S.T. } \sum_k x_{ik} = 1 \text{ for all } i \tag{2-2}$$

$$\sum_k x_{kk} = p \tag{2-3}$$

$$x_{ik} \in \{0,1\} \text{ for all } i \text{ and } k \tag{2-4}$$

$$x_{ij} \le x_{jj} \text{ for all } i \text{ and } j \tag{2-5}$$

$$\sum_m x_{ijkm} = x_{ik} \text{ for all } i, j, k \tag{2-6}$$

$$\sum_k x_{ijkm} = x_{jm} \text{ for all } i, j, m \tag{2-7}$$

$$x_{ijkm} \in \{0,1\} \text{ for all } i, j, k, m. \tag{2-8}$$

[119] See O'Kelly (1987), pp.393-404.

[120] See Campbell (1994b), pp.387-405.

[121] See Skorin-Kapov et al. (1996), pp.582-593.

[122] See Ernst/ Krishnamoorthy (1996), pp.139-154.

[123] See Ebery (2001), pp.447-458.

[124] See Skorin-Kapov et al. (1996), pp.582-593.

The *p*-hub median problem is *NP*-hard[125]. Moreover, even if the locations of the hubs are fixed, the allocation part of the problem is proved to be *NP*-hard by Kara[126].

***P*-hub center problem**

The *p*-hub median problem can sometimes lead to unsatisfactory results when worst-case O-D distances are excessively large. In order to avoid this drawback, *p*-hub center problems provide one option. It is a minimax problem. Campbell was the first to formulate and discuss the p -hub center problem. He defined three different types of problems.[127]

(I) The maximum cost (or time) for any O-D pair is minimized.

(II) The maximum cost (or time) for movement on any single link (origin-to-hub, hub-to-hub and hub-to-destination) is minimized.

(III) The maximum cost (or time) of movement between a hub and an origin (or a destination) is minimized.

The first type is applied to a hub system involving perishable or time sensitive items. The second type involves items that require some preserving/processing such as heating or cooling which is available at the hub locations. The third type is applied to the cases, in which feeder transportation is subject to a time limit.

Kara and Tansel provided a combinatorial formulation of the single-allocation *p*-hub center problem[128]. Ernst et al. developed a new two-index formulation for this problem[129]. Compare with that by Kara and Tansel, Ernst's formulation has more continuous variables but fewer constraints and requires less CPU time.

Ernst et al[130] defined a new variable r_k as the maximum collection/distribution cost (or time) between hub *k* and the nodes that are allocated to hub *k*. *Z* is a free variable to represent the objective. The objective is to minimize the maximum of the costs (or time) between any pair of nodes *i* and *j*. With previously defined parameters and decision variables, the formulation based on the first definition of the problem is expressed as:

$$Min \;\; Z \tag{2-9}$$

S.T. $$r_k \geq c_{ik} x_{ik} \quad \text{for all } i, k \tag{2-10}$$

$$Z \geq r_k + r_m + \alpha c_{km} \quad \text{for all } k, m \tag{2-11}$$

$$r_k \geq 0 \quad \text{for all } k \tag{2-12}$$

Also constraints (2-2)-(2-5).

[125] See e.g. Alumur/ Kara (2008), p.5; Campbell (1996), p.926.

[126] See Kara (1999).

[127] See Campbell (1994b), pp.387-405.

[128] See Kara/ Tansel (2000), pp.648-655.

[129] See Ernst et al. (2009), pp.2230-2241.

[130] See Ernst et al (2009), pp.2230-2241.

Kara and Tansel [131] have proved that it is *NP*-complete by a reduction from the dominating set problem. Ernst et al. [132]studied the allocation sub-problem of the single allocation *p*-hub center problem when hub locations are fixed. They also proved the *NP*-hardness of this problem.

Hub set covering problem

The hub set covering problem is to locate least number of hubs to cover all demand. In FLPs, demand nodes are considered to be covered if they are within a specified distance of a facility that can serve their demand. Campbell defined three coverage criteria for HLPs just as for the p-hub center problem[133]. The O-D pair (i, j) is covered by hubs k and m if

(I) the cost from i to j via k and m does not exceed a specified value,

(II) the cost for each link in the path from i to j via k and m does not exceed a specified value, and

(III) each of the origin-hub and hub-destination links meets separate specified values.

Campbell[134] presented the first mixed integer formulations for hub set covering problem. Kara and Tansel[135] presented and compared three different linearization of the original quadratic model and presented a new linear model. Ernst et al.[136] strengthened the formulation of Kara and Tansel by replacing a constraint with its aggregate form. This formulation performs better in terms of CPU time requirement than the formulation by Kara and Tansel. Wagner[137] proposed new formulations for both single and multiple allocation hub set covering problems. By his proposed preprocessing techniques he rules out some hub assignments and thus the formulations require less number of variables and constraints than that of Kara and Tansel's formulation. We introduce the hub set covering problem with the formulation by Ernst for the sake of brevity.

$$Min \sum_k x_{kk} \tag{2-13}$$

$$\text{S.T. } r_k + r_m + \alpha c_{km} \le \beta \text{ for all } k, m \tag{2-14}$$

s.t. (2-2), (2-4),(2-5), (2-10), and (2-12)

where β is the cover radius.

Kara and Tansel[138] proved that the single allocation hub set-covering problem is *NP*-hard.

[131] See Kara/ Tansel (2000), pp.648-655.

[132] See Ernst et al (2009), pp.2230-2241.

[133] See Campbell (1994b), pp.387-405.

[134] See Campbell (1994b), pp.387-405.

[135] See Kara/ Tansel (2003), pp.59-64.

[136] See Ernst et al (2005).

[137] See Wagner (2007), pp.932-938.

[138] See Kara/ Tansel (2003), pp.59-64. All in the recent researches. See Alumur/Kara (2008), pp.9-11 and p.14.

Hub location problem with fixed cost

Hub location problem with fixed cost discussed in this paper is discussed by some former studies in two sub-problems, i.e. uncapacitated hub location problem (UHLP) and capacitated hub location problem (CHLP)[139].

O'Kelly[140] introduced the single-allocation hub location problem with fixed costs by adding fixed costs of opening hubs into the *p*-hub median problem and making the number of hubs as a decision variable. He formulated the problem as a quadratic integer program. Campbell presented the first linear programming formulations for multiple/single-allocation uncapacitated/capacitated hub location problems[141]. Other formulations were also proposed by Abdinnour-Helm and Venkataramanan[142], Ernst and Krishnamoorthy[143].

The hub location problem with fixed cost can be formulated as follows.

$$Min \sum_i \sum_j \sum_k \sum_m w_{ij} x_{ijkm} (c_{ik} + c_{mj} + \alpha c_{km}) + \sum_k F_k x_{kk} \tag{2-15}$$

S.T. (2-2), (2-4)-(2-8)

where F_k is the fixed cost of opening a hub at node *k*.

When the locations of the hubs are fixed, the allocation sub-problem is the same as the allocation sub-problem of the *p*-hub median problem, which has been proved to be *NP*-hard by Kara[144]. So the hub location problem with fixed cost is also *NP*-hard.

Actually, all of the four classical HLPs discussed above are *NP*-hard (except for some special cases[145]). Thus the exact solution potential for these problems is limited.

2.2.5. Conventional assumptions and corresponding extensions

As we go through conventional models in early studies, we find three strict assumptions for model simplification that are extensively cited in later researches[146]:

(1) Hubs are fully interconnected with direct links[147];

---

[139] See the pertinent literature review by Hekmatfar/ Pishvaee (2009), pp. 243-270.

[140] See O'Kelly (1992), pp. 293–306.

[141] See Campbell (1994b), pp.387-405.

[142] See Abdinnour-Helm/ Venkataramanan (1998), pp.31-50.

[143] See Ernst/ Krishnamoorthy (1999), pp.141-159.

[144] See Kara (1999).

[145] For example, Sohn and Park have proved that the single allocation p-hub median problem in a two-hub system has a polynomial time algorithm. They also showed that it is NP-hard as soon as the number of hubs is three. See Sohn/ Park (2000), pp.17-25.

[146] For example in *p*-HLP by O'Kelly (1987), pp.393-404, in multi-allocation *p*-hub median location model by Campbell(1996), pp.923-925, in *p*-hub median location problem with fixed costs by O'Kelly (1992), pp.292–306, in *p*-hub center location problem by Campbell (1994), pp. 387–405. These three assumptions were summarized by Alumur/ Kara (2008a), p.2.

(2) EOS on inter-hub link is incorporated in models with a fixed discount factor α [148] ;

(3) There is no direct link between any non-hub nodes, namely strict and restrictive H/S network.

As we can see, these assumptions are so strict and simplified that they can hardly comply with reality. Some researchers have made efforts to relax one or more of these assumptions and incorporate more realistic characteristics into hub location models.

Extension 1: incompletely connected backbone network

One relaxation to the conventional assumptions is that hubs are not completely interconnected[149]. This relaxation was extended as hub arc location problems that seek to locate certain number of discounted hub arcs to minimize the total transportation cost[150].

This assumption was also relaxed by introducing fixed or setup cost of inter-hub links in the network. The cost may depend on the length of the links or the geographical location of the links. If links are possible with different capacities, the setup cost may vary accordingly. When fixed cost of inter-hub links is considered, it may result in incompletely connected backbone network when other constraints are not violated. However, when fixed cost of link is included, the corresponding transportation cost is no longer linear but concave.

Extension 2: variable discount rate on links

Conventional H/S network models also make some simplifications on the cost structure of links. It is assumed that the discount rate on inter-hub links is exogenously fixed, i.e. discount rate on inter-hub links is independent of flow. It is also assumed that only flow on inter-hub links is discounted. These assumptions can simplify the modeling for the pure ground H/S network without much loss of reality, since inter-hub links are always served by trucks with higher efficiency. However, these assumptions can hardly be applied to air-ground H/S networks, since the air cost rate on inter-hub links is much higher than that on feeder links. Moreover, sometimes cost rate on both backbone and feeder links depends on flow[151]. It should be pointed out that the assumption of flow-independent cost function not only miscalculates the total network cost but also may erroneously select hubs and make suboptimal allocation decisions[152]. More rational transportation cost calculation may be with a cost function that rewards the higher volume with lower cost rate.[153]

Extension 3: nonrestrictive and nonstrict H/S network configuration

[147] This implies that p hubs are connected by $p(p-1)/2$ undirected hub links or $p(p-1)$ directed hub links.

[148] See O'Kelly/ Miller (1994), p.31; Skorin-Kapov et al (1996), pp.582-593; Ernst/ Krishnamoorthy (1996), pp.139-154.

[149] It was first studied by Chou. See Chou (1990), p.247.

[150] See Campbell et al. 2005(a), pp.1540–1555; Campbell et al. (2005b), pp.1556–1571; Campbell et al (2003), pp.555–574.

[151] See Horner/ O'Kelly (2001), p.255.

[152] See Repolho et al. (2010), p.957.

[153] That means the inter-hub link cost function is concave. Therefore, total network cost is minimized by forcing some interacting pairs to use non–least-cost path. Passenger inconvenience (in terms of travel time) makes the network with the flow-dependent cost function inappropriate for passenger air network. However, this gives freight network the opportunity to maximize load factors and EOS regardless of routing. See O'Kelly (1998), p.610.

In restrictive and strict H/S networks, all non-hub nodes are directly connected to hub(s), i.e. there is no direct link between any non-hub nodes[154]. With less links, this network configuration has the advantage of contracture simplicity, high resource utilization and low setup costs. In many applications, however, this rigid restriction may be undesirable concerning longer service time. In the case of passenger air network, for example, most passengers prefer more convenient direct service, and would not make a stopover at a third airport. In EDS network, this restriction may lead to losing important niche market to rivals because of the inability to offer direct service. Sometimes it is even not economical to make transshipment through hub(s). Actually, when EOS by aggregating cannot compensate the detour, it is more economical to delivery directly.

Thereby one variation of the restrictive and strict H/S network allows direct links between non hubs so that channeling flows through hubs is not required but adopted if the cost is lower and time constraints are not violated. Another variation is to include stopovers or tours in tributary networks. Models including non-hub to non-hub direct links or/and stopovers on feeder routes find that they always incur lower transportation cost and require fewer feeder vehicles with higher load factors than their strict counterparts[155]. Fig.2-8 shows an example of nonrestrictive and nonstrict H/S network with incompletely connected backbone network. When parcels travel from Hub A to Hub C, it must be transshipped at either Hub B or Hub D. Demand node 1 and 2 forms a tributary route with one stopover. Demand nodes 3, 4, 5 and 6 form a ring route, which only needs one feeder truck. Demand 7 serves as a regional consolidation center with a tree-formed sub-network. It is also connected directly with Demand node 6. Demand node 11 is multi-allocated to both Hub A and D.

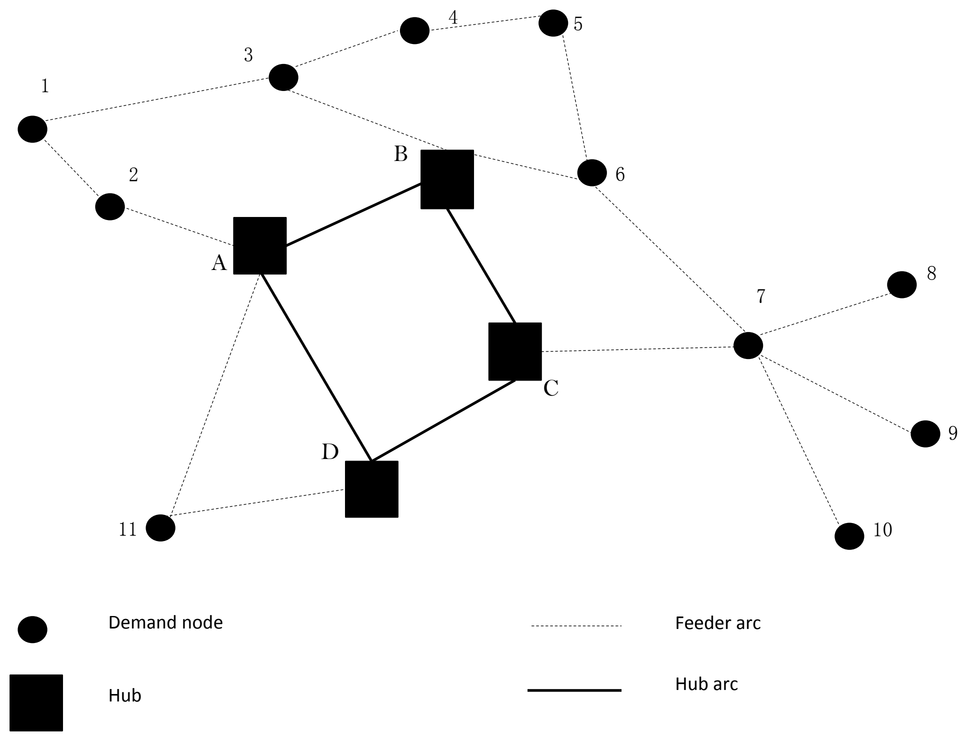

Figure 2-8: Example of nonrestrictive and nonstrict H/S network

[154] See O'Kelly/ Miller (1994), p.32; Bryan/ O'Kelly (1999), p.276; Zäpfel/Wasner (2002), p.208.

[155] See Aykin (1995), p.217; Kuby /Gray (1993), p11.

2.3. Current research gaps and research focuses of this dissertation

2.3.1. Current research gaps

- Multimodal transportation networks for EDS

Multimodal transport (also known as combined transport) is "transport of goods under a single contract, but is performed with at least two different means of transport. The carrier is liable (in a legal sense) for the entire service, even though the service is performed by several different modes of transport (by rail, sea and road, for example)".[156] Under the background of globalization and a highly competitive market environment, multimodal transportation networks are integral to EDS providers who offer time-definite service over a large area. Carriers have to decide which transport mode to select so that all delivery tasks can be fulfilled in specified time window with minimum cost. Air transport implies fast delivery over long distance but at high cost, while ground transport has the opposite implication. In addition, trucking allows more flexibility in scheduling.

The research topic "intermodal transport" shares some similarity with "multimodal transport" in our research. Intermodal transport is defined in 1993 by the European Conference of Ministers of Transport (ECMT) as "the carriage of goods by at least two different modes of transport in the same loading unit (an Intermodal Transport Unit or ITU) without stuffing or stripping operations when changing modes. It emphasizes the integrated use of two or more modes of transportation to deliver goods from origin to destination in a seamless flow."[157] However, practically most of the route is traveled by rail, inland waterway or ocean-going vessel and with the shortest possible initial and final journeys by road.[158] In academic research it also has different emphasis, such as seamless transshipment, mode connectivity costs[159], transit delays[160] and mode choice[161]. However, researches under these two topics are based on similar H/S network structures, in which hubs (consolidation terminals, rail yards, intermodal platforms, and so on) serve as mode connection point.[162] So researches on intermodal networks can also be referred by researches on multimodal transportation networks for EDS.

However, after we search papers on both of these two topics, i.e. "intermodal transport" and "multimodal transport", we find that OR has focused mostly on transport problems of uni-modal transport modes. The number of studies on both topics is very limited. In network planning for EDS or postal service that adopt multimodal transport, researchers first divide the whole network into several sub-networks according to transport modes and then study the sub-network(s) individually and separately[163]. The shortage of researches

[156] See Macharis/Bontekoning (2004), p.400-416. It also includes classification and application of this problem.

[157] See Crainic et al (2007), p. 468.

[158] See Macharis/ Bontekoning (2004), p.400.

[159] See Groothedde et al. (2005), pp.567–583.

[160] See Ziliaskopoulos/Wardell (2000), pp. 486–502.

[161] See e.g. Kreutzberger (2008), pp.973-993; O'Kelly/ Lao (1991) , pp.283-297; McGinnis (1989), pp. 36-46.

[162] See Macharis/ Bontekoning (2004), p.408.

[163] See e.g. Grünert / Sebastian (2000), pp.289-309; Grünert/ Sebastian/Thärigen (1999), p.16 ; Büdenbender/ Grüner/ Sebastian (2000), pp.364–380; Armacost et al. (2004), p.15; Kuby / Gray (1995), pp.1-12.

on multimodal network planning, especially those for EDS, may result from both theoretical and practical reasons.

Since 1990 a substantial number of analytical publications specifically addressing intermodal transport issues have appeared. However, the use of OR in intermodal transport papers were still very limited until about 10 years ago[164]. Theoretically, the planning for intermodal freight transport is more complicated than that for uni-modal systems. It involves at least two modes, which have their own specific characteristics in perspective of infrastructure and transport vehicles.

EDSI itself is an emerging industry[165] and the corresponding academic studies are less than those for traditional industries. In practice, EDS is often partially performed by sub-carriers (in legal language it is referred as "actual carriers")[166], except when some large EDS providers own the whole service network by themselves. Carriers focus on one transport mode and outsource the left to other service providers, who have their own resource advantages. Or sometimes EDS providers have local agencies that are responsible for their own profits and losses, so that the carriers have difficulty or no impetus to reorganize the whole network. Moreover, the requirement to plan multimodal transportation networks is not strong. Multimodal networks are only necessary for large countries or regions. At the same time those areas should be relatively developed so that the requirement for premium EDS is strong and intensive.[167]

- Strategic planning for EDS networks

Compared with studies on tactical planning, studies on strategic planning for EDS or postal service networks are relatively less. Hubs are always predetermined so that most studies are conducted at the tactical planning level[168]. The main reason may be that HLP is a relatively new ramification of FLP, which has a long history. Fruitful research results of FLPs can shed light on researches on HLPs[169]. Another reason may be that strategic planning covers a quite long horizon, while tactical and operational planning is in demand more often.

- Large-scale HLPs

As we have mentioned in Sec.2.1.2, problems from real-life application are always of large scale. Although techniques for OR have been developed and new heuristics have been developed, due to the problem type, size and complexity, most researches on HLPs with OR are based on the CAB data set with 25 nodes. Some are based on the AP data set with 50 nodes and few are with 100 nodes[170] and 200 nodes[171]. The largest instances

[164] Researches on network models for terminal location decisions include those by e.g. Ishfaq/Sox (2011), pp.213-230; Groothedde/ Tavasszy (1999), pp.43-57; Van Duin/ Van Ham (1998), pp.11-14.

[165] See Sec.1.1.2 for the development of the EDSI.

[166] See Macharis/ Bontekoning (2004), p.414.

[167] Even for countries like Turkey, we cannot find research on multimodal network planning for nationwide EDS until 2010. See Çetiner et al (2010), pp.109–124.

[168] See e.g. Grünert/ Sebastian(2009), pp. 289-309; Lin/Chen (2004), pp. 271–283; Kim et al. (1999), pp. 391-407.

[169] For readers who are interested in FLPs, please refer to recent literature reviews, e.g. Liu (2009), and also books. See Liu (2009), pp.157-165; Klose/ Drexl (2005); Drezner/Hamacher (2001); Domschke (1997).

[170] See Labbe/ Yaman (2008), pp.19-33.

that are involved in OR, to the best of our knowledge, are those by Resende and Werneck, who proposed a heuristic combining fast local search and path-relinking within a multi-start heuristic for the uncapciated facility location problem and test its performance under instances with 1000 nodes.[172] Wager proposed an exact cluster solution procedure for a cluster hub location problem (capacitated hub location problem under a non-restrictive policy) and provided optimal results for instances with maximum 500 nodes.[173]

2.3.2. Research focuses of this dissertation

This project-based dissertation is applicable to strategically rebuilding or modifying the current network to offer trans-city overnight EDS in most part of China with a large-scale, multi-modal and time-definite network.

First, the network planning is conducted at a strategic level. We simultaneously determine the hub location, demand allocation and air service on hub arcs. The first two are involved in HLPs. The last one, which is actually aircraft fleet ownership decision, will be embedded in our hub network planning problem, since it has a quite long planning horizon. However, service network design problems (SNDP), such as routing and scheduling, are not included in this dissertation[174].

Second, the planning is for an air-ground multimodal network. China is a quite large country. The network planned in this dissertation supports nationwide trans-city overnight EDS[175] so that an air-ground multimodal transportation network is a must. In order to simplify the problem, we predefine the network with a fully interconnected/star H/S structure. The backbone network is served by air, while the feeder networks are served by trucks. We consider important aspects of the multimodal transportation network, such as transshipment time, location of connection points, overall transportation cost and overall delivery time.

Third, we plan a network for time-definite EDS. Delivery time has become a critical competitiveness for EDS providers and, therefore, is an important consideration in the network planning. Pertinent researches on this topic include "the latest arrival HLPs"[176] and "hub set covering models"[177]. However, both of them take the total delivery time as coverage criterion[178], resulting in different coverage radius for different hubs. In this dissertation, we define coverage constraints for both tributary and backbone networks[179]. There are two reasons. For one thing, we roughly presume a fully interconnected backbone network during the strategic planning. Thus, we define a quite loose time window for the air network so that in practice there is still cost-

[171] See Contreras et al (2011), pp.41-55; In the case for Swiss parcel delivery service, the demand nodes in the models are about 200.

[172] See Resende/ Werneck (2006), pp.54-68.

[173] See Wagner (2007), pp.391-401.

[174] SNDPs require the determination of a set of routes for the assets and a set of shipments on the asset network, that satisfy all customer demand with an acceptable level of service at minimum cost. and without violating capacities of service legs. See Barnhart et al. (2002), p. 239.

[175] For readers who are interested in pickup and delivery systems in metropolis, please refer to Hall (1996), pp.173-187; Hall (2001), pp.331-338.

[176] See e.g. Yaman (2007), pp. 906-919; Kara/ Tansel (2001), pp.1408-1420; Wagner (2004), pp.1751-1755.

[177] See e.g. Tan/ Kara (2007), pp.28-39; Kara/ Tansel (2003), pp.59-64; Alumur/Kara (2008), pp.1349-1359.

[178] Three well-acknowledged coverage criteria were defined by Campbell. See Campbell (1994b).pp. 387-405. Also see Sec.2.2.4.

[179] Since we assume that the time window for backbone air network is enough for direct flight between any of the potential hubs, we just omit these constraints in the model formulation.

saving opportunity by transshipment at a third airport. The corresponding air routing problem will be studied in tactical planning. For another, in practice the to-be-planned network also supports the Same Day EDS within each hub region, which makes the definition of the maximal hub coverage radius necessary. We plan the network for the whole service system from the top by considering the two key services in a nationwide scope. As the Same Day EDS only involves feeder networks, it is not included in this dissertation.

Fourth, we base our planning on the current network by considering the current facilities with "Sunk Cost Theory" and aircraft fleet with flow-dependent cost function. We study two extension models, one with the optimal aircraft fleet composition and the other one under the constraints of the current fleet composition.

Finally, our research is aim at strategically planning a large-scale network with 281 demand nodes. With regard to this, we resort to hybrid GAs to get good solutions in bearable time but without the guarantee of finding optimal solutions. We try to improve the performance of the algorithm with customized improvement techniques for different procedures of GAs. We also evaluate their performance with public test data sets.

The models studied in this dissertation are a combination of hub location problem with fixed cost and hub set covering problem which have been introduced in Sec.2.2.4. Since neither of the two problems addresses both of the two issues :(1) the time constraints and (2) the total cost, especially transportation cost. In particular, the hub location problem with fixed cost, on the one hand, decides the optimal hub number, hub location and allocation of demand nodes to hubs with the objective of minimizing the hub fixed cost and transportation cost. But it does not account for time constraints. The hub set covering problem, on the other hand, takes the coverage criterion as constraints, while neglecting transportation cost, which is an important consideration for EDS providers.

In this dissertation we combine these two problems together to complement each other**.** We minimize both hub fixed cost and transportation cost under the constraints that all demand nodes are covered by their "home" hub, by assuming that the time window for the air network is feasible for direct flight between any potential hubs. Actually we are not the first one to make this combination. The initial work, to our knowledge, is proposed by Sim[180], who named it as Hub Covering Flow Problem (HCFP). The author studied the best configuration of H/S network that minimizes the total cost of opening hubs and transportation cost, while satisfying a maximum flying distance constraint between the hub and non-hub airports. The author provided two formulations for the HCFP and compared the results by using the 50-node Australia Post (AP) data set and commercial solver Xpress-MP. Later Campbell[181] included the coverage constraints into the multi-allocation *p*-hub median model and the hub arc location model to reflect the situations encountered during his work with the trucking industry. However, these are the only two works we can find that study a similar model as ours.

By combining the hub set covering problem and hub location problem with fixed cost, we first propose the basic model, which conforms to the three conventional assumptions in Sec.2.2.5, i.e. fully interconnected hubs, fixed discount rate on hub arcs and no direct link between non-hub nodes. Then we extend the basic model by

[180] See Sim (2007), available on internet: http://ir.uiowa.edu/etd/124.

[181] See Campbell (2009), pp. 3107-3116.

eliminating the assumption of fixed discount rate on hub arcs. Air service selection problem for the backbone air network is included by considering a cost select function that can be easily transformed into a piecewise linear cost function. We consider two different situations- whether the air service selection is subject to the current fleet composition (*Ext.2*) or not (*Ext.1*) (see Fig.2-9).

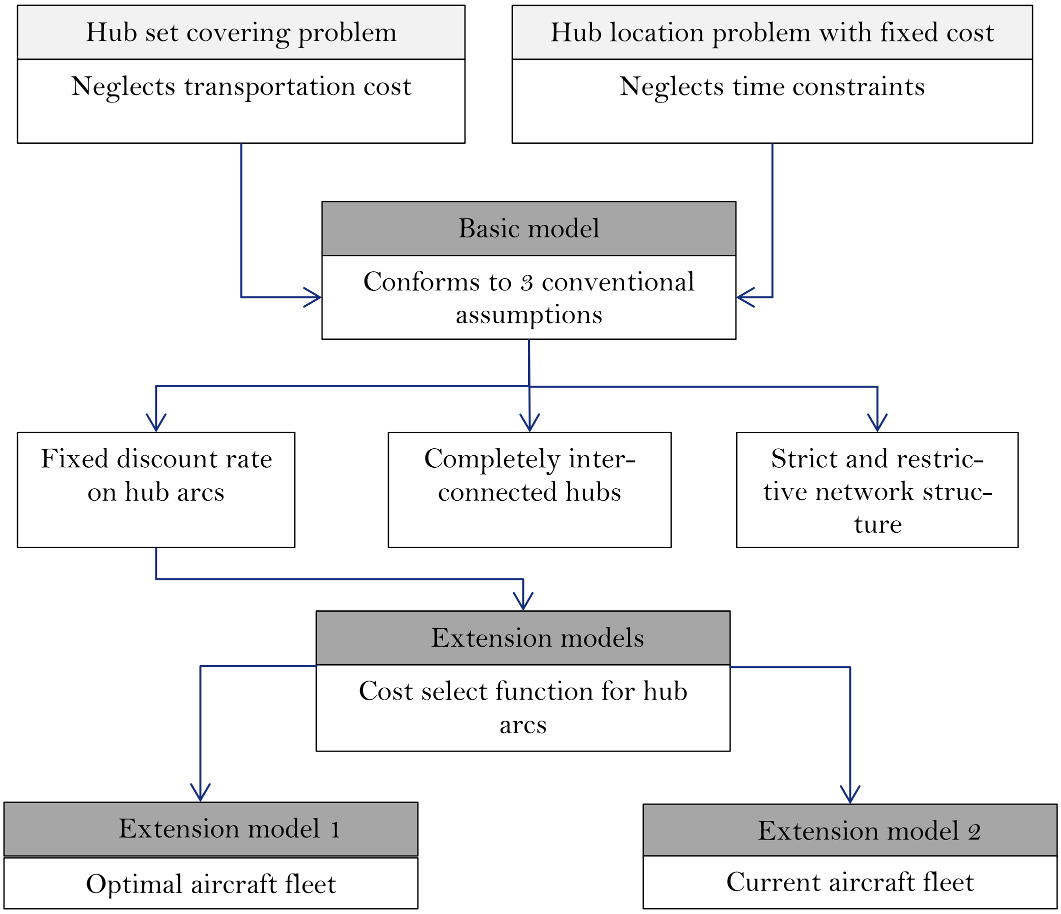

Figure 2-9: The basic model and its extensions

3. Model formulation

3.1. Network and service description

We plan a transportation network to support nationwide trans-city overnight EDS, which is defined as key service in the service system[182]. Any order placed by the end of one business day (e.g. before 6:00 P.M.) should be accomplished in the morning (e.g. before12:00 A.M.) of the next business day or next business day (e.g. before 6:00 P.M.). This service will be offered in target market defined by Company A, including most of the economically developed cities in China[183]. In this section, we describe in detail the network structure, parcel paths and service on the target market.

3.1.1. Network structure

In order to provide a nationwide trans-city overnight EDS, we resort to a multimodal H/S network. We define it with a fully interconnected/star shaped H/S structure for the strategic planning in this dissertation. Particularly, several cities in the potential hub set are chosen as hubs, while all the other cities belonging to the target market are allocated to one of these hubs subject to the maximum direct distance constraints. All the hubs also serve as gateways for the air network, which are fully interconnected by direct air service. The non-hub cities are connected to their "home" hubs by direct ground service[184].

Single allocation criterion is adopted here under management and cost considerations. This decision is made by the management of Company A to maintain a non-overlapping organization structure and clear responsibility for subsidiaries. Furthermore, additional feeder links also increase fixed cost of the network.

Every hub is per se a demand node, which is also called in-hub demand node in this dissertation. Other demand nodes that are not chosen as hubs are also called normal demand nodes. Service area of a hub, i.e. all the cities allocated to it together with the hub itself, is defined as a hub region. Every normal demand node is equipped with a city station, in which all parcels from and to that city are consolidated and sorted (called local sorting). Every hub is equipped with a regional ground consolidation center, in which all parcels from or to that region are consolidated and sorted (called regional sorting). It also serves as a gateway of the air network for all the demands from or to other hub regions to make air-ground transshipment. Because it also consolidates and sorts parcels from and to that in-hub demand node, the city station and the regional consolidation center in the hub city is geographically coincident.

The network structure is illustrated in Fig.3-1. The dashed line encloses ground the tributary/feeder networks, while the bold line encloses the air backbone network. Ground feeder links are represented by the fine dashed arrows and air backbone links are represented by the bold arrows.

---

[182] See Tab.1-2 in Sec. 1.3.3.

[183] See Sec.6.1.1.

[184] As we have mentioned in Sec.2.2.5, hubs can be connected by air routes with stopovers if time permits. Moreover, several cities in one hub region can be served by one truck route with several stopovers. However, routing problems are not included in this dissertation. The network is simplified with fully interconnected/star H/S structure.

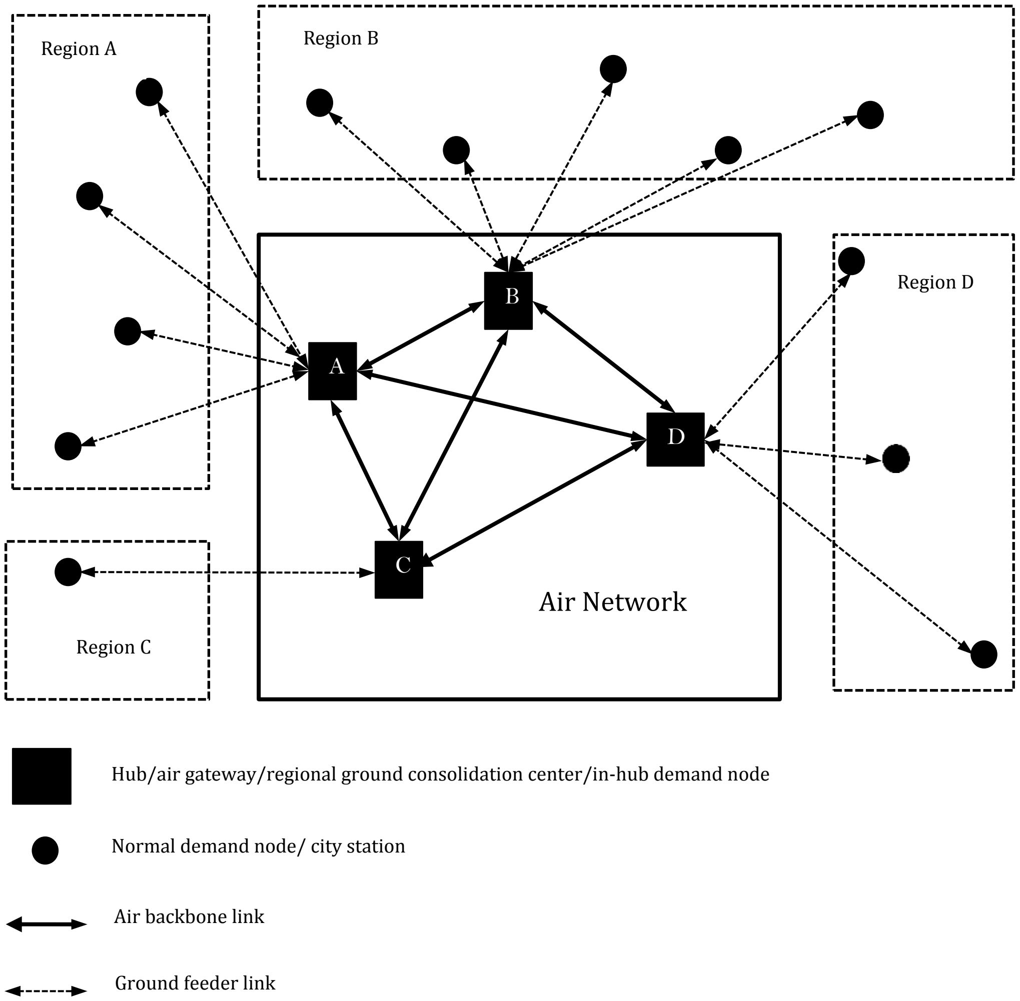

Figure 3-1: Multimodal H/S network for the overnight EDS

3.1.2. Parcel paths[185]

The EDS begins from orders online or through service hotline from customers. A fleet of vans responds to the requests before the local cutoff time in the late afternoon and picks up the parcels from shippers and deliver to the city station. After twilight local sorting, parcels with destination in other cities are loaded on a truck and transported from the city station to the consolidation center in the hub city through highway. However, this feeder transportation to consolidation center can be saved for a hub city, since the city station and the consolidation center is identical so that the twilight local sorting is implemented in the consolidation center. That also means the cut-off time for an in-hub demand node can be later than other normal demand nodes in that hub region. After twilight regional sorting in the consolidation center, parcels with destination in other re-

[185] We differentiate several related concepts in this dissertation. An "arc" or a "link" is a node-to-node direct connection by vehicles or aircraft. A "path" is defined as a sequence of arcs used to deliver a parcel from its origin to its destination. A "route" is defined as a sequence of arcs traveled by the same vehicle. In this dissertation all the arcs, paths and routes are directed, i.e. in only one direction.

gions are transported by direct flights to destination hubs. Parcels with destination in the same region are left in the consolidation center. Early in the morning of the next business day, after all trans-regional parcels arrive at their destination hub by air, both trans-regional and intra-regional parcels are sorted in the consolidation center according to their destinations, called sunrise regional sorting. A fleet of trucks is then dispatched from the regional consolidation center to every city station in that region. After the parcels arrive at their destination cities, city stations unload the parcels from trucks, conduct sunrise local sorting and reload them onto a fleet of vans for city distribution. Likewise, feeder transportation to city station is saved if the destination city is that hub city. In this respect, after sunrise regional sorting parcels with destination in this hub city go directly into sunrise local sorting procedure.

Different nodes along a parcel path play different roles, which are described in Tab.3-1. Although hubs and in-hub demand nodes are geographically identical, they play different roles in the network. So we consider them as two different types of nodes in the network planning.

Nodes		**Roles in parcel path**	**Attributes**
Hub		Origin Hub	Cut-off time: the latest time that feeder trucks arrive at the hub, the earliest time that twilight regional sorting starts
		Destination Hub	Set-up time: the latest time that aircraft arrive at the hub, the earliest time that sunrise regional sorting starts
Demand node	Normal demand node	Origin city	Cut-off time: the latest time that parcels arrive at city station, the earliest time that twilight local sorting starts
		Destination city	Set-up time: the latest time that feeder trucks arrive at city station, the earliest time that sunrise local sorting starts
	In-hub demand node	Origin city	Cut-off time: the latest time that parcels arrive at regional consolidation center, the earliest time that twilight local sorting starts (later than the cut-off time for normal demand node)
		Destination city	Set-up time: the latest time that sunrise regional sorting finishes, the earliest time that sunrise local sorting starts (earlier than the set-up time for normal demand node)

Table 3-1: Nodes along parcel path and their attributes

In this regard, there are 4 different kinds of paths for trans-regional parcels, i.e. normal demand node to normal demand node, normal demand node to in-hub demand node, in-hub demand node to normal demand node and in-hub demand node to in-hub demand node. Without the stretch of backbone air transportation, intra-regional parcel paths are quite similar to those for trans-regional. Or we can regard it as a degenerated hub arc in intra-regional parcel paths. Consequently, totally there are 8 different parcel paths, 4 for trans-regional parcels and 4 for intra-regional ones (see Fig.3-2). Fig.3-3 and Fig.3-4 illustrate in detail the parcel paths from a normal demand node to a normal demand node and from an in-hub demand node to an in-hub demand node, respectively.

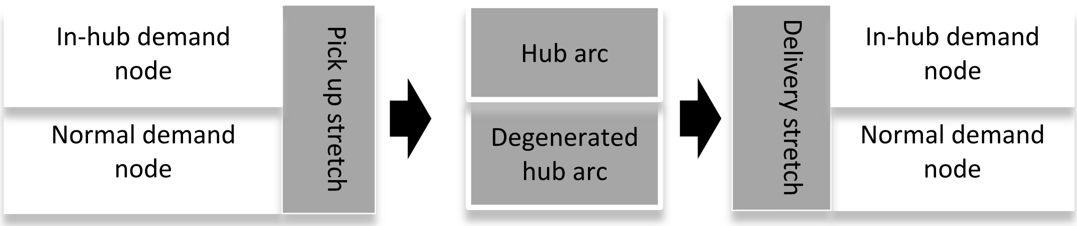

Figure 3-2: Description of parcel paths

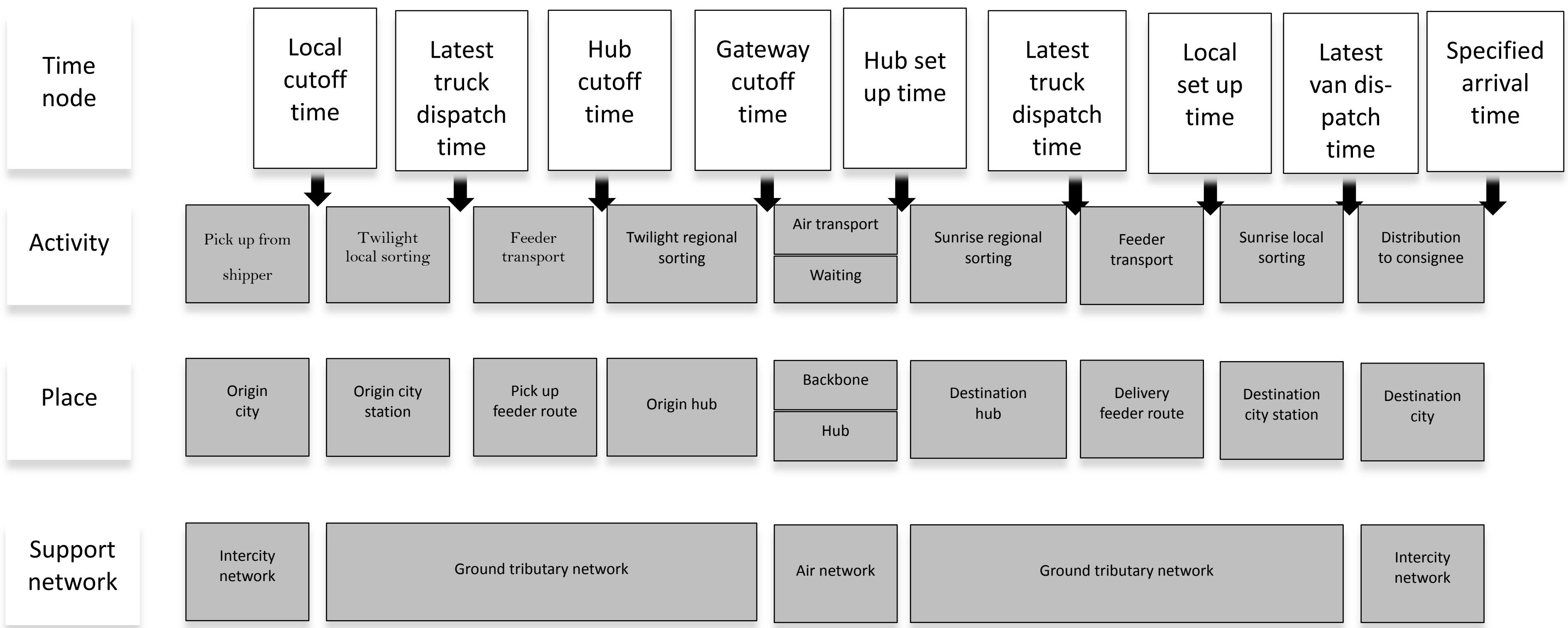

Figure 3-3: Parcel path from a normal demand node to a normal demand node

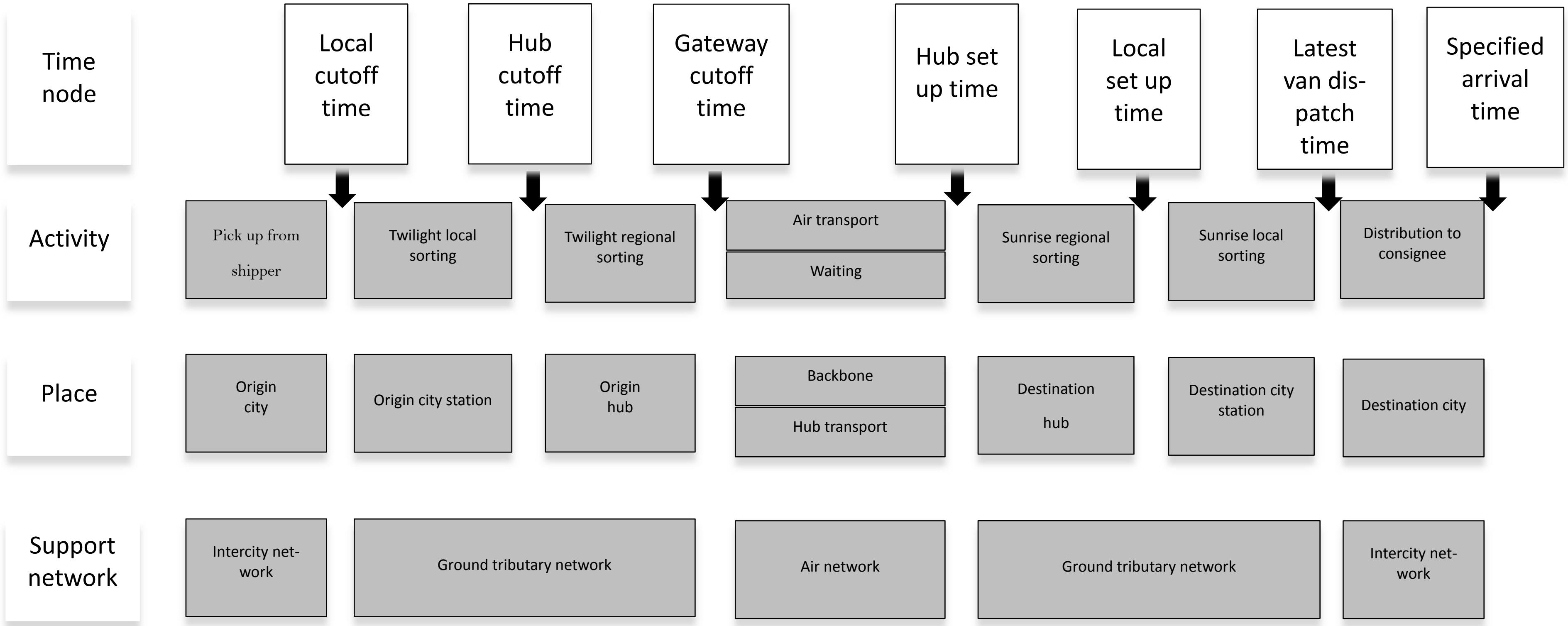

Figure 3-4: Parcel path from an in-hub demand node to an in-hub demand node

3.1.3. Service specifications in the target market

According to the market strategy of the company, the key services will be offered to the target market. But actually there are two kind of key services, i.e. next morning and next day EDS[186]. In other words, the service quality in terms of delivery time for each city in the target market depends on its role in the network, namely hub or non-hub city. Tab.3-2 specifies the services in the target market according to the origin and destination of delivery orders. Hub A and Hub B represent different hub nodes or in-hub demand nodes, while Region A and Region B are different regions covered by Hub A and Hub B. The first column stands for origins, while the first row stands for destinations. We can check the service policy according to the origin and destination of a delivery order. As we can see from the table, the arrival time depends on whether the destination is hub or not, no matter the delivery order is an intra-regional or a trans-regional one. Hub cities are provided with EDS with shorter delivery time, i.e. later cutoff time and earlier arrival time than other non-hub cities, since feeder ground transportation by truck is saved.

O-D	**Hub A**	**Non-hub in region A**	**Hub B**	**Non-hub in region B**
Hub A	next morning[187]	next day	next morning	next day
Non-hub in region A	next morning	next day	next morning	next day
Hub B	next morning	next day	next morning	next day
Non-hub in region B	next morning	next day	next morning	next day

Table 3-2: Service specifications in the target market

3.2. Basic model

3.2.1. Model assumptions

In the basic model, two decisions are made simultaneously to minimize the overall cost: (1) the location of the hubs and (2) the allocation of non-hub nodes to "home" hubs. We follow the three conventional assumptions in HLPs listed in Sec.2.2.5: (1) a fully interconnected backbone network, (2) a fixed discount factor α on hub arcs and (3) a strict and restrictive H/S network structure. Other assumptions are listed as follows. Most of them are also applicable to the extension models. If relaxed, they are pointed out accordingly.

(1) All non-hub cities are directly connected to their unique "home" hubs. That is, the tributary network is "star" shaped with the single-allocation criterion.
(2) Hubs are fully linked by direct flight. That is, the backbone network is fully interconnected.
(3) The time window for the earliest departure and the latest arrival of aircraft is the same for all potential hubs.
(4) Air cost is linearly dependent on distance and traffic volume.
(5) There are no capacity constraints on hubs or arcs.

[186] For next morning EDS, parcels arrive before 12:00 next business day. For next day EDS, parcels arrive before 18:00 next business day. Also see Tab.1-2 in Sec.1.3.3.

[187] It is actually intra-city EDS, which is not studied in this dissertation.

(6) Demand volume between all O-D pairs is deterministic.
(7) Time window for air transportation satisfies the longest direct flight between hubs.
(8) All the distance applies to the triangle inequality.

It is also to underline that there is no assumption on symmetric demand volume, which is an assumption quite commonly applied by previous studies[188].

3.2.2. Model formulation

In this section, we propose a 0-1 integer programming model for our network planning problem. HLPs can be formulated in a variety of ways, depending on the form of decision variables. Different formulation methods for the same problem result in different numbers of variables and constraints[189]. In this dissertation, we use the four-subscript formulation method introduced by Campbell for a linear model[190].

The model is based on a network, in which the demand node set N and the potential hub set H ($H \in N$) are identified. The model is to locate hubs in the potential hub set H and allocate the remaining demand nodes in N to the located hubs under the constraints of the maximum hub coverage radius with the objective of minimizing the total cost. The hub coverage radius is considered in the form of distance by transforming the time window for the feeder transportation into a distance bound with an average speed of truck on highway. The decision variables and parameters for the basic model are listed in Tab.3-3 and 3-4.

Decision variables	**Description**
x_{ik} $(i \in N, k \in H)$	Refer to both location and allocation variable. $x_{ik} = 1$ if node i is allocated to hub k, otherwise 0. When $x_{kk} = 1$, k is a hub.
y_{ijkm} $(i, j \in N, k, m \in H)$	Refer to path variable. $y_{ijkm} = 1$, if flow from node i to j via hubs k and m in that order. Otherwise 0. Note $y_{ijkm} = x_{ik} \bullet x_{jm}$.

Table 3-3: Decision variables in the basic model

Parameters	**Description**

[188] See e.g. Kuby /Gray (1993), pp. 1-12; Lin et al. (2003), pp.255-265.

[189] Campbell et al (2005b) listed three different integer programming formulations appeared in former studies: binary allocation variable by O'Kelly, four-subscript formulation by Campbell (1994) and flow tracking variable by Ernst and Krishnamoorthy (1996, 1998a). The first one is for single allocation problems. Because of the quadratic terms in the objective function, it was proved to be a poor choice for solution. The second and third ones are applicable to multiple-allocation problems. See Campbell et al.(2005b), p.1557; O'Kelly (1987), pp. 394-404; Campbell (1994), pp. 387-405; Ernst/Krishnamoorthy (1996), pp.139-154; Ernst/Krishnamoorthy (1998a), pp.100-112.

[190] See Campbell (1994), pp. 387-405. It is also to be pointed out that the linear model is at the cost of more variables and constraints.

N	Set of demand nodes
H	Set of potential hubs, $H \subseteq N$
w_{ij}	Demand volume from node $i(i \in N)$ to node $j(j \in N)$
fh_k	Fixed cost of opening a new hub or expanding an existing hub at potential hub node $k(k \in H)$
β_{km}	Cost rate by air in backbone network (per kilo- kilometer), $k, m \in H$
γ_{ik}	Cost rate by truck in feeder network (per kilo- kilometer), $i \in N$, $k \in H$
d_{ik}	Distance by highway between non-hub node $i(i \in N)$ and its "home" hub $k(k \in H)$
d_{km}	Distance by air from hub node k to m, $k, m \in H$, $d_{km} = d_{mk}$
D_k	Distance bound as hub coverage radius of potential hub node k, $k \in H$

Table 3-4: Parameters in the basic model

With the defined parameters and decision variables, we formulate the objective function as follows:

$$\text{Minimize} \sum_{k \in H} fh_k x_{kk} + \sum_{i \in N} \sum_{j \in N} \sum_{k \in H} \sum_{m \in H} w_{ij} y_{ijkm} \left(\gamma_{ik} d_{ik} + \beta_{km} d_{km} + \gamma_{mj} d_{mj} \right) \tag{3-1}$$

In the objective function, the first term sums the fixed costs of hubs either for new establishment or expansion. The second term calculates the total transportation cost. The demand volume and costs are calculated on daily basis.

$$\text{S. T.} \sum_{k \in H} x_{ik} = 1 \quad \forall i \in N \tag{3-2}$$

$$x_{ik} \le x_{kk} \quad \forall i \in N, \forall k \in H \tag{3-3}$$

$$x_{ik} \in \{0,1\} \quad \forall i \in N, \forall k \in H \tag{3-4}$$

$$0 \le d_{ik} \cdot x_{ik} \le D_k \quad \forall i \in N, \forall k \in H \tag{3-5}$$

$$\sum_{m \in H} y_{ijkm} = x_{ik} \quad \forall i, j \in N, \forall k \in H \tag{3-6}$$

$$\sum_{k \in H} y_{ijkm} = x_{jm} \quad \forall i, j \in N, \forall m \in H \tag{3-7}$$

$$\sum_{m \in H} \sum_{k \in H} y_{ijkm} = 1 \quad \forall i, j \in N \tag{3-8}$$

$$y_{ijkm} \in \{0,1\} \quad \forall i, j \in N, \forall k, m \in H \tag{3-9}$$

Constraints (3-2) and (3-4) ensure that every non-hub node is allocated to exactly one hub in conformity to the single-allocation criterion. Constraints (3-3) state that non-hub node cannot be allocated to another node unless that node is a hub. And when a node is hub itself, it is allocated to itself. Constraints (3-5) make sure that all non-hub nodes are allocated to their "home" hubs subject to the distance criteria. Constraints (3-6) state that if node i is allocated to hub k, all the demand from node i to any other node j must go through some hub m. If node i and j are in the same hub region, $m=k$. Constraints (3-7) have a similar interpretation involving the demand from any node i to node j, which is assigned to hub m. Constraints (3-8) guarantee that there is one path for each O-D pair. Note that constraints (3-8), together with constraints (3-6) and (3-7), ensure that every node is allocated to only one hub. Constraints (3-9) are the constraints of path variables.

In the basic model, when the hub locations are fixed, the allocation sub-problem is the same as the allocation sub-problem of the single allocation p-hub median problem, which has been proved to be *NP*-hard by Kara[191]. So the basic model here belongs to the class of *NP*-hard problems. If we set $n = |N|$ and $h = |H|$, the model has $(n^2h^2 + nh)$ binary variables and $(2n^2h + n^2 + 2nh + n)$ linear constraints.

3.3. Flow-dependent air cost

3.3.1. HLPs based on current network

Virtually the vast majority of HLPs planned at the strategic level are based on the premise that the network is planned from scratch. That is to say, there is no need for planners to consider facilities and other fixed assets, such as vehicles and equipment, in current networks. However, in many real-life cases the planning is based on current networks that may need to be expanded, merged or even shrunk[192], i.e. modification or reconstruction of current networks.

This is the pragmatic problem that is faced by many EDS providers. In our case the to-be planned network is partially owned by a relatively large EDS provider in China and is at present supporting its nationwide express delivery business. Besides the resources that the company shares with its partner, a lot of proprietary assets have been invested in the current service network. Specifically, due to capacity bottleneck and market competition, it has built dozens of regional ground hubs, consolidation centers and air gateways consecutively without globally planning the network in the past few years. Built specially for express delivery business, these facilities are invested by Company A alone and not shared by its partner. Moreover, it owns several aircraft exclusively for EDS. We must take these resources into account, when we globally plan the new network.

- **Current facilities**

Current facilities can sometimes be considered in the network planning with conditional facility location problems or facility relocation problems. Conditional facility location problem, whose name is acknowledged in literatures, actually deals with facility expansion problems. It tries to find the best location for p new facilities, when some existing facilities are already located in the area. Customers are assumed to get service from the closest facility whether existing or new, so most probably they are conditional p-center problem or conditional p-median problem[193]. Since they are dependent on the number of existing facilities, q, they are also called the conditional (p, q)-median/center problems[194]. Facility relocation problem handles with both facility expansion

[191] See Kara (1999). Sohn and Park also proved it NP-hard when the hub number is larger than 2. See Sohn/ Park (2000), pp.17-25.

[192] See e.g. ReVelle (2007), pp.533-540.

[193] See Tamir (2005), p.50.

[194] See e.g. Drezner (1995), p.525.

and phase-out problems. As a matter of fact, it is taken up as a reactive strategy for the organization to adjust itself in time to reality. This problem is also discussed with dynamic location methods[195].

In this dissertation we resort to the "Sunk Cost Theory" to consider current facilities with static models. We regard the value of the facilities that cannot be transferred into the new network or cannot be retrieved by transferring on the market as sunk cost. In the model we only consider extension cost of the potential hub nodes that are currently equipped with consolidation centers or regional ground hubs.

- **Current aircraft**

Aircraft is another kind of important fixed assets for EDS providers. Compared with truck and van, it incurs much higher purchase price and maintenance cost and has less chance to change hands on the market. Acquisition, possession, mothballing or selling of an aircraft is a relatively long-term decision compared to decision of other ground vehicles and equipment.

As is anticipated, more hubs will be installed in the new network, indicating that more flight routes are necessary. Actually, it is unnecessary and also uneconomical to satisfy all the air freight demand by self-owned aircraft due to the small volume on most of the inter-links. Company A intends to continue to adopt a mixed air freight service strategy by out-sourcing and self-owned aircraft with the objective of minimizing the air cost. That means air freight tasks are fulfilled by both self-provided and commercial air services. Therefore, the strategic planning of the EDS network is also faced up with the following issues: how to assign current aircraft in the new air network; is it more economical to purchase new aircraft and what type to choose or stop using current aircraft.

The air service selection decision, basically speaking, aircraft fleet ownership decision, is seldom included in HLPs. However, in perspective of management the decision on whether outsourcing or self-provided service has much longer planning horizon than other tactical decision, such as vehicle routing and scheduling problems. For this reason, we include this decision in our network planning. The rest of this chapter is dedicated to research on air service selection decision and aircraft ownership decision by distinguishing cost functions of different air freight services[196]. In other words, the air cost is no longer as simple as that in the basic model. We separate different air services with flow-dependent cost functions, including service from self-owned aircraft, normal suppliers on air freight market and contracted suppliers with quantity discount rate. The objective is to minimize the total cost by optimally determining the aircraft fleet ownership (*Ext.1*) or by making full use of the current self-owned aircraft (*Ext.2*). Meanwhile, we also investigate if air service selection decisions can in turn impact on hub location decisions.

Incorporating flow-dependent air cost functions into HLPs will break the one of the three classical assumptions[197], i.e. fixed discount rate on hub arc, which means that the average cost rate of inter-hub links is fixed or

[195] For example, Melachrinoudis and Min determined the optimal timing of relocation and phase-out in the planning horizon using a dynamic, multiple objective, and mixed-integer programming model. Wang et al. studied a budget constrained location problem in which they simultaneously consider opening some new facilities and closing some existing facilities. See Melachrinoudis/Min (2000), pp.1-15; Wang et al. (2003), pp. 2047-2069.

[196] This method is also applicable to tributary network with ground transportation.

[197] For details, please refer to Sec.2.2.5.

independent on the flow and it is also lower than that on feeder route. In our case, this simple but arbitrary assumption will seriously distort the problem.

For one thing, the travel cost is not always proportional to the flow. It is obvious that the average air cost decreases with the increase of the flow, when we consider link fixed cost. Moreover, contracted air freight suppliers always offer discount to the order that is above the predetermined minimum quantity (or threshold).

For another, backbone cost rate is not always lower than feeder cost rate. The classical assumption is consistent with the fact that travel cost in ground backbone network is lower than that in feeder network due to the possibility to utilize larger vehicles by bundling parcel streams in backbone transportation[198]. Economies of scale (EOS) typically result in lower cost rate on backbone links in comparison to that on tributary links. Higher cost in tributary networks may be the reason why nearly all HLRPs include vehicle routing problems (VRPs) for tributary networks rather than for backbone networks. However, the case in air-ground networks for EDS seems to be on the opposite side. Globalization and economic development propel the multimodal transportation systems to offer faster and seamless service. In EDSI, parcels and mails are consolidated for the sake of faster transportation mode rather than only for the sake of lower transportation cost. Backbone transport is more often than not accomplished by air with higher cost rate rather than by more economical lorry. Therefore, the backbone network, a vital factor of the cost and delivery time for the EDS, must be paid more attention.

For these reasons, we put more emphasis on the backbone network and study service selection problem on backbone network rather than on tributary networks.

3.3.2. Review on flow-dependent cost function

As we have illustrated in the last section, exogenously determined and fixed travel cost rates oversimplify the problem and sometime run counter to real-life situations. In most real-life cases, transportation cost rates depend on the flow, when the fixed cost of the links is under consideration. In particular, average cost decreases with the increase of the flow, or the cost (per kilometer) increases at a decreasing rate when the flow increases. In mathematical terminology, the cost function is concave rather than linear as commonly assumed.[199] The mathematical definition of a concave function is:

$$\frac{d\Omega}{df} > 0 \tag{3-10}$$

$$\frac{d^2\Omega}{df^2} < 0 \tag{3-11}$$

where Ω denotes the cost in per kilometer and f denotes the inter-hub flow.

[198] Former studies estimated that backbone ground transportation cost for parcel delivery service takes up about 15%-25% of the total cost, while pickup and delivery cost takes up 35–60%. See Wasner/Zäpfel (2004), p.406; Salhi/ Rand (1989), pp.150–156.

[199] See e.g. Ben-Ayed (2012), p.7; Kimms (2005), p.301.

Fig.3-5 shows typical examples of total cost (per kilometer) of concave cost function and linear cost function, respectively. Fig.3-6 shows examples of average cost (per kilometer*kilo). The dashed line in both figures represents concave cost function, while the solid line represents the linear cost function assumed in traditional HLPs. The linear cost function in this dissertation is used to represent the cost function of ordinary service from the air freight market, while the concave cost function represents the situations when the fixed cost of self-owned vehicle is taken into account or when discount with minimum order quantity is offered by contracted suppliers.

Fig.3-5 demonstrates clearly the error resulting from the oversimplification of the discount rate by the classical assumption. The point where the two lines intersect represents the minimum flow (or threshold) to achieve the discount rate in traditional models. When the inter-hub flow is less than this threshold, the discount rate in traditional models is too large so that the transportation cost is underestimated. When the flow is larger than this threshold, the constant discount rate is not enough to reward the volume. Both of these two situations distort the backbone link cost. In addition, a constant cost rate can in no way encourage flow agglomeration through inter-hub links so that the network does not make full use of the benefit from EOS. What is even more serious is that the oversimplification of the discount rate may lead to wrong decisions on hub location and demand allocation, since the location and allocation decisions are largely dependent on discount factors[200].

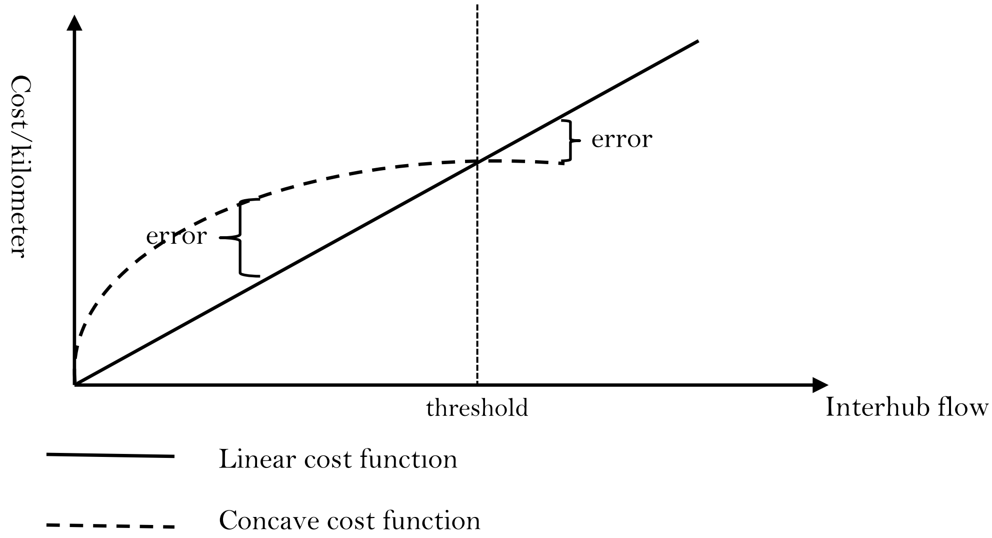

Figure 3-5: Total cost of concave cost function and linear cost function

(Source: O'Kelly/Bryan (1998), p.607)

[200] See O'Kelly /Bryan (1998), pp.605-616.

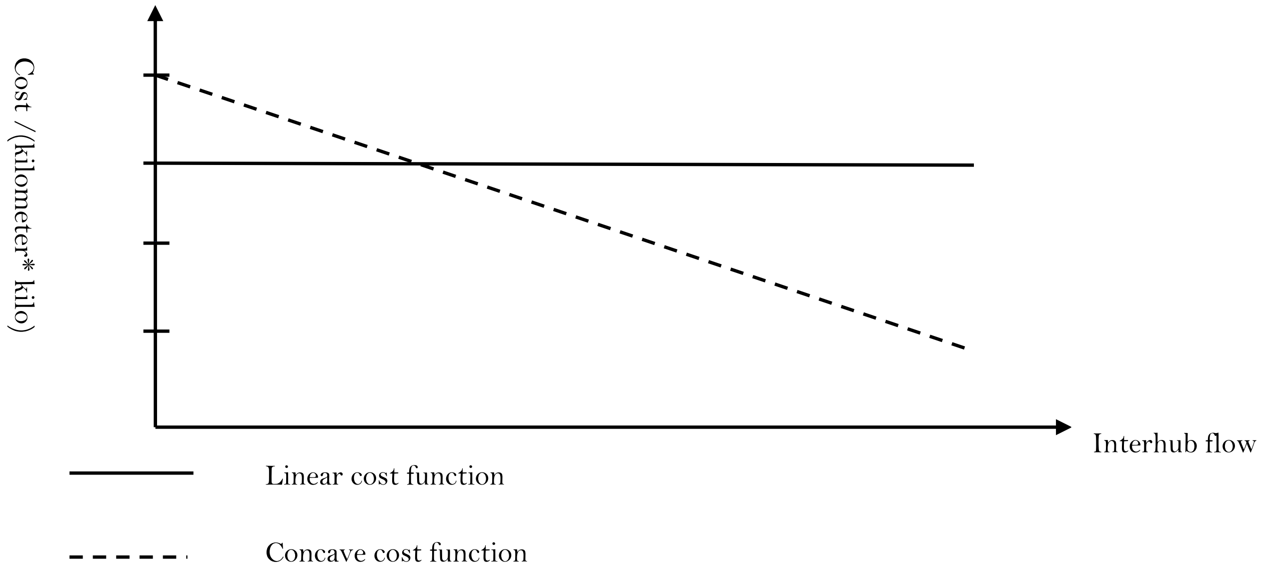

Figure 3-6: Average cost of concave cost function and linear cost function

(Source: O'Kelly/Bryan (1998), p.607)

In light of these considerations, transportation cost function that is more close to reality is called for by HLPs. By noting that applying a discount rate on inter-hub links while disregarding the flows contradicts the intention of EOS in H/S networks, some researchers introduced different approaches to model the transportation cost rate in H/S networks, including flow-dependent cost rate[201] and flow threshold-based cost rate[202]. It is also believed that besides inter-hub links, spokes (links between hub and non-hub) can also bring about EOS with sufficient flow[203].

O'Kelly and Bryan[204] set up a model called FLOWLOC, which is dedicated to modeling EOS on inter-hub links. The nonlinear cost function monotonically increases with the flow. The nonlinear function is then approximated by a piecewise linear function in such a way that the lower envelope of this piecewise linear function approximates the nonlinear function. The authors solved the hub location model with 20 nodes and 2-4 hubs to optimum by using a standard mixed integer programming package. In their computational tests, each non-linear cost function was approximated by only two pieces of linear function. However, a single instance required as much as several days of computational time. Later, Klincewicz proposed a location enumeration procedure that is based upon TS and GRASP for the linearized FLOWLOC model. When the hub location is fixed, the remaining multi-allocation problem can be solved as an uncapacitated facility location problem (UFLP). [205]

[201] See e.g. Bryan (1998), pp.315–330; O'Kelly/Bryan (1998), pp.605–616; Horner/ O'Kelly (2001), pp.255-265.

[202] See e.g. Podnar/ Skorin-Kapo (2003), pp.207-228; Podnar et al (2002), pp.371-386; Campell (1994b), pp. 387-405.

[203] See Kimms (2005), pp.293-317; Racunicam/ Wynter (2005), pp. 453–477; Horner/ O'Kelly (2001), pp.255-265.

[204] See O'Kelly/ Bryan (1998), pp.605-616.

[205] See Klincewicz (2002), pp.107-122.

Racunicam and Wynter also defined a nonlinear concave cost function. Based on research results on polyhedral properties of this problem, a linearization procedure along with two variable-reduction heuristics was developed. [206] It has been proved by Hamacher et al. that when polyhedral properties of the linear hub location model are adopted, the cuts in the mixed-integer programming are quite efficient at removing non-integer solutions.[207]

Besides linearization of nonlinear cost functions, the most commonly used technique to handle concave cost function is approximating it with a piecewise linear function so as to allow the use of linear programming solvers, just as O'Kelly and Bryan[208] did. In this respect, the network planning with concave cost is actually the minimum concave-cost network flow problem (MCNFP), which is also called concave piecewise linear network flow problem (CPLNFP) in some papers[209].

However, when the cost function is piecewise linear, the search for the optimal solution might be still very difficult, since this problem possesses high dimension of feasible region in network and thus a large number of local optima.[210] A reference of the exact solution of MCNFP defined over polytope is that of Horst and Tuy[211]. In this book, numerous versions are presented for partitioning or covering the polytope of feasible solutions by polyhedral cones, and then solving one-dimensional problems over each cone. The method works much like a branch and bound (B&B) algorithm by enumerating wisely the extreme points that are likely to provide a global optimal solution of the problem. Consequently, it is inappropriate to solve large-scale MCNFP. More recent references for exact algorithms include variable disaggregation by Croxton et al.[212] and bilinear relaxation-based algorithm by Nahapetyan and Pardalos[213]. Other works on exact solutions include those by Balakrishnan and Graves[214], Verter and Dincer[215], Cominetti and Ortega[216] and Kim and Pardalos[217]. While exact methods can be only applied to small-scale problems, heuristics are for large-scale MCNFPs. An important contribution in this respect is that from Minoux[218], who presented a path flow exchange algorithm based on repeated calculations of shortest path and shifts of flow across the shortest paths. This algorithm, which has the additional benefit of being very easy to implement, is still widely used today by many MCNFPs.

---

[206] See Racunicam/ Wynter (2005), pp. 453–477.

[207] See Hamacher et al (2004), pp. 104-116.

[208] See O'Kelly/ Bryan (1998), pp.605-616.

[209] See e.g. Kim/ Pardalos (2000), pp.225-234; Croxton et al (2007), pp.146-157; Nahapetyan/ Pardalos (2007), pp.71-91;

[210] See Nahapetyan/ Pardalos (2007), p.72.

[211] See Horst/ Tuy (1996).

[212] See Croxton et al. (2003), pp.1268-1273; Croxton et al. (2007), pp.146-157.

[213] See Nahapetyan/ Pardalos (2007), pp.71-91;

[214] See Balakrishnan/ Graves (1989), pp. 175–202.

[215] See Verter/ Dincer (1995), .pp.1141–1160.

[216] See Cominetti/ Ortega (1997).

[217] See Kim/ Pardalos (2000), pp.216–222.

[218] See Minoux (1989), pp.313-360.

3.3.3. Cost select function

As we have mentioned in Sec.3.3.2, several methods have been adopted to model transportation cost in H/S networks, including piecewise linear cost function, nonlinear concave cost function and threshold-based discount cost rate. In our opinion, the economic explanation for applying nonlinear concave cost function is somewhat weak, at least in our case, since the parameters in the function can hardly find economic interpretation in reality. Meanwhile, the threshold-based discount cost rate still depends on an exogenously fixed discount rate α . The solid line in Fig.3-7 presents a typical threshold-based cost function commonly applied in many studies. As a matter of fact, the threshold-based cost function has an unreasonable gap at the threshold point; the bold line implies that the cost is higher with less flow than the cost with more flow in certain interval on the right side of the threshold point.

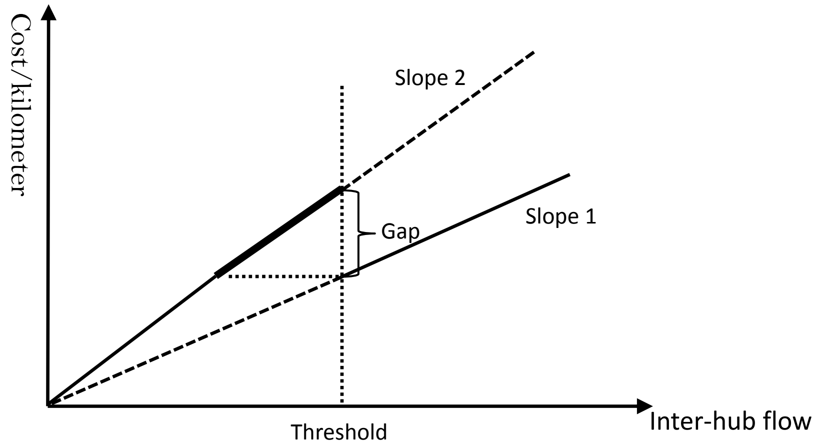

Figure 3-7: Threshold-based discount cost rate

Parameters	Description
P	set of services
k_p	fixed cost of service $p(p \in P)$ (per kilometer)
χ_p	variable cost of service $p(p \in P)$ (per kilo*kilometer)
u_p^l	lower bound of feasible domain for service $p(p \in P)$
u_p^u	upper bound of feasible domain for service $p(p \in P)$

Table 3-5: Parameters for the cost function

In this dissertation, we employ a cost select function that can be calculated easily into a piecewise linear function if all cost parameters are defined. We specify the parameters in the cost select function in Tab.3-5. We use p to index the linear cost functions that represent different services in set P. We define service with a specific cost structure as one service, such as service with normal cost rate, service with quantity discount rate and services by different types of self-owned aircraft. The cost function for each one kind of service can be described with an intercept k_p (the fixed cost), a slope χ_p (the variable cost), a lower bound of feasible domain

u_p^l and an upper bound u_p^u. Air service selection problem determines the service type for each hub link by minimizing the air cost with cost select function $SF_{kl}(w)$ (see Eq.3-12).

$$SF_{kl}(w) = \min \begin{cases} d_{kl}(\chi_1 w + k_1) & u_1^l \le w \le u_1^u \\ d_{kl}(\chi_2 w + k_2) & u_2^l \le w \le u_2^u \\ d_{kl}(\chi_3 w + k_3) & u_3^l \le w \le u_3^u \\ \vdots & \vdots \\ \vdots & \vdots \\ d_{kl}(\chi_P w + k_P) & u_p^l \le w \le u_p^u \end{cases} \quad p \in \{1, 2, \cdots, P\} \quad k, l \in H \tag{3-12}$$

In some studies a specific cost function with variable cost and fixed cost is defined for each link[219]. In this dissertation, we separate the distance factor from the cost rate so that the travel cost for each link can be estimated with the same cost function for the sake of simplicity. But these two methods share the common concern that the variable cost depends on the volume while the fixed cost dose not.

However, actually the fixed cost is not completely proportional to the flight distance due to the considerable taking-off and landing cost incurred for every flight and other costs such as crew cost. Some works consider fixed cost based on time, e.g. annual fixed cost or daily fixed cost[220]. It may be more accurate to adopt daily fixed cost in this dissertation if the aircraft fleet is exclusive for the target EDS, i.e. key services, and one aircraft conduct one flight every day. However, in our case, when a self-owned aircraft finishes the job of overnight EDS, it will be dispatched for other business if possible. Moreover, in real-life instances aircraft may follow routes with one stopover if time permits. For these reasons, we have to fairly allocate the fixed cost to the overnight express delivery business. So considering the fixed cost based on distance is more reasonable for our case.

When all the cost parameters in Eq.3-12 are defined, the cost select function can be easily transformed into a piecewise linear cost function, which is the lower envelope of all the linear cost functions, such as the solid line in Fig.3-8. For example, when the cost select function $SF_{kl}(w)$ is defined as Eq.3-13, the corresponding piecewise linear function $PW_{kl}(w)$ is expressed as Eq.3-14 and is illustrated in Fig.3-8. As we can see, the first turning point (0.5)of the piecewise linear cost function is determined by the intersection of slope 1 and slope 2, while the second intersection (1) is determined by the upper capacity of slope 2.

$$SF_{kl}(w) = \min \begin{cases} d_{kl} 5w & 0 \le w \\ d_{kl}(3w+1) & 0 \le w \le 1 \\ d_{kl}(w+4) & 0 \le w \le 2 \end{cases} \quad k, l \in H \tag{3-13}$$

$$PW_{kl}(w) = \begin{cases} d_{kl} 5w & 0 \le w \le 0.5 \\ d_{kl}(3w+1) & 0.5 < w \le 1 \\ d_{kl}(w+4) & 1 \le w \le 2 \end{cases} \quad k, l \in H \tag{3-14}$$

[219] See e.g. Kimms (2005), p.301; O'Kelly (1998), p.609.

[220] See e.g. Ben-Ayed (2012), p.7.

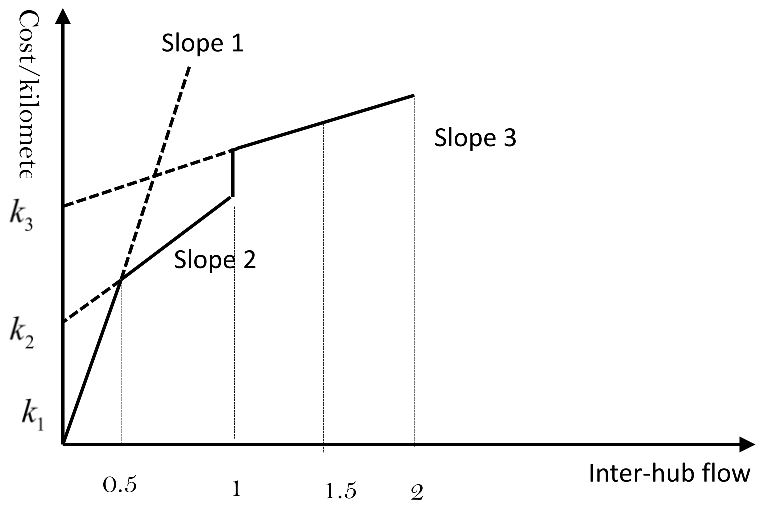

Figure 3-8: Example for piecewise linear cost function

Note that when parameters χ_p and k_p in Eq.3-12 satisfy the following inequalities, the corresponding piece-wise linear cost function is a concave one, i.e. unit cost is monotonically increasing at a decreasing rate.

$$\chi_1 \geq \chi_2 \geq \ldots \geq \chi_P > 0 \tag{3-15}$$

$$k_1 \leq k_2 \leq \ldots \leq k_P \tag{3-16}$$

3.3.3.2 Smooth treatment of the cost select function

The piecewise linear cost function described in Fig.3-8 is relatively perfect that after each threshold the cost function comes to a more economical one. But cost select functions under real-life instances are most often not so perfect (see e.g. Fig.3-9).

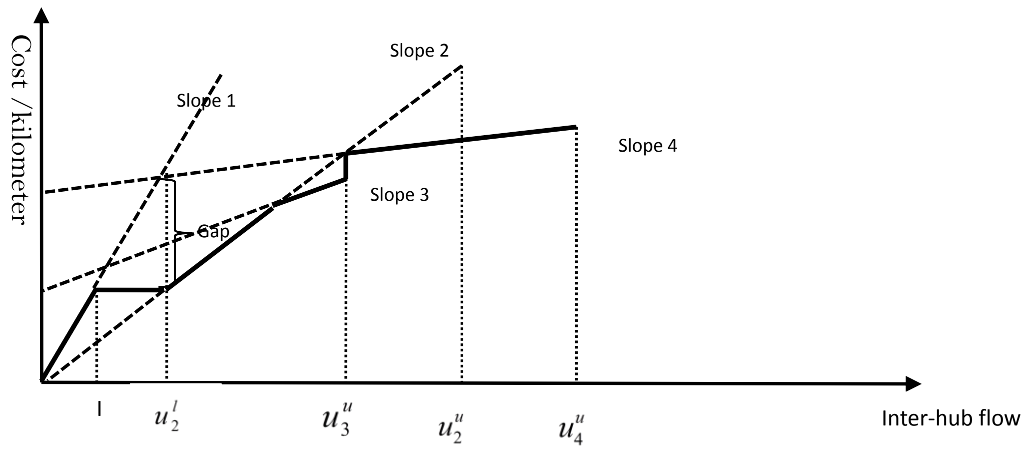

Figure 3-9: Normal piecewise linear cost function with constraints on feasible domain

Slope 1 and Slope 2 represent normal freight rate and quantity discount rate with minimum quantity of u_2^l respectively. Consequently, $k_1 = k_2 = 0$, $u_1^l = 0$ and $u_1^u = u_2^u = \infty$. Slope 3 and 4 represent air services by self-owned aircraft with fixed cost k_3 and k_4, variable cost χ_3 and χ_4, minimum quantity $u_3^l = u_4^l = 0$ and maximum capacity constraints u_3^u and u_4^u. Owing to the minimum quantity for the discount rate required by service providers, the first cost function jumps down to the second one with a gap described in Fig.3-9. But in reality service providers often fairly charge the same price when the flow lies in the interval of $\left[\mathrm{I},\ \mathrm{u}_2^l\right]$. In this respect, we modify the cost select function into a more rational one with the horizontal solid line in the interval $\left[\mathrm{I},\ \mathrm{u}_2^l\right]$ (see the horizontal bold line Fig.3-9).

Therefore, we introduce a parameter λ to smooth the cost select function and make up such unreasonable gap by dividing each former cost function into two segments if necessary and redefining lower and upper bounds (see Eq.3-18, 3-19 and Eq. 3-20) for the cost select function. The modified cost select function $SF'_{kl}(w)$ is defined as Eq.3-17).

$$SF_{kl}^{'}(w) = \min \begin{cases} d_{kl}(\chi_1 w + k_1) & \lambda_{11} \le w \le \lambda_{12} \\ d_{kl}(\chi_1 \lambda_{12} + k_1) & \lambda_{12} < w \le \lambda_{13} \\ d_{kl}(\chi_2 w + k_2) & \lambda_{21} < w \le \lambda_{22} \\ d_{kl}(\chi_2 \lambda_{22} + k_2) & \lambda_{22} < w \le \lambda_{23} \\ \vdots & \vdots \\ \vdots & \vdots \\ d_{kl}(\chi_p w + k_p) & \lambda_{p1} < w \le \lambda_{p2} \\ d_{kl}(\chi_P \lambda_{p2} + k_P) & \lambda_{p2} < w \le \lambda_{P3} \end{cases} \quad p \in \{1, 2, \cdots, P\}\ k, l \in H \tag{3-17}$$

where

$$\lambda_{p1} = u_p^l \tag{3-18}$$

$$\lambda_{p2} = u_{p+1}^l \tag{3-19}$$

$$\lambda_{p3} = u_p^u \tag{3-20}$$

The modified cost select function has several advantages. First, the cost select function is less artificial and can be economically explained better than those nonlinear ones in our case. Every parameter in the cost select function has a corresponding economical meaning and can be estimated accordingly. Second, when all the cost parameters are determined, the cost select function can be easily transformed into a piecewise linear cost function, in which the cost rate is endogenously determined by the flow and the turning points of the cost rate are determined mutually by all parameters. Third, only one kind of service is allowed on each link, which is required by the management out of operational reasons. So combination of different services is forbidden even when it is more economical. Fourth, the cost select function is in a generalized form, which can simultaneously include as many cost functions as possible for self-owned vehicles and outsourced service with different quantity discount policies. Finally, the cost select function undergoes a special smooth treatment to generate more reasonable cost rate. This method can be easily applied to other flow-dependent cost function formulation, e.g. threshold-based cost rate.

3.4. Extension models

3.4.1. Extension model 1

The basic model assumes that the air cost for all inter-hub links is proportional to the traffic volume. Extension model 1 (*Ext.1*) is to minimize the total cost of the network by considering different cost functions for different service types but without numerical constraints on these service types. The modified cost select function for inter-hub links proposed in last section is incorporated into the basic model to determine the hub location, demand allocation, and most importantly here, the optimal aircraft fleet composition.

We continue to apply all assumptions of the basic model in Section 3.2.1 except the 4th one, which is about the inter-hub link cost function. We list the additional assumptions for *Ext.1* in the following.

(1) Commercial air freight service is uncapacitated. In other words, there are no constraints on the upper bounds.

(2) The cost select function is the same for all inter-hub links.

(3) All direct flights satisfy the time window for the air network.

We define two more decision variables for *Ext.1* (see Tab.3-6). They are also applied in *Ext.2*. M is a very large constant.

Decisions	Description
f_{km}	Flow on inter-hub link *(k,m)* , $k,m \in H$
z_{km}^{p}	$z_{km}^{p} = 1$ if service $p\,(p \in P)$ is applied on hub link (k,m), otherwise 0.

Table 3-6: Decision variables for extension models

Ext.1 is formulated as follows.

Minimize $$\sum_{k \in H} fh_k x_{kk} + \sum_{i \in N}\sum_{j \in N}\sum_{k \in H}\sum_{m \in H} w_{ij} y_{ijkm} (\gamma_{ik} d_{ik} + \gamma_{mj} d_{mj}) + \sum_{k \in H}\sum_{m \in H}\sum_{p \in P} SF'_{km}(f_{km}) z_{km}^{p} \tag{3-21}$$

In the objective function Eq.3-20, the first term represents the hub fixed cost. The second and third terms represent the tributary and backbone transportation cost respectively. The constraints of the model are as follows.

S.T. $$f_{km} = \sum_{i \in N} \sum_{j \in N} y_{ijkm} w_{ij} \quad \forall k,m \in H \tag{3-22}$$

$$u_p^l z_{km}^p \le f_{km} \; \forall k,m \in H, \; k \ne m, \; p \in \{1,2\ldots,P\} \tag{3-23}$$

$$f_{km} \le u_p^u + (1 - z_{km}^p) M \;\; \forall k,m \in H, \; k \ne m, \; p \in \{1,2\ldots,P\} \tag{3-24}$$

$$\frac{\sum_{i \in N}\sum_{j \in N} y_{ijkm}}{n^2} \le \sum_{p \in P} z_{km}^p \le 1 \quad \forall k,m \in H, \; k \ne m, n = |N| \tag{3-25}$$

$$z_{km}^{p} \leq x_{kk} \quad \forall k,m \in H\,,\ k \neq m\,,\ p \in \{1,2\ldots,P\} \tag{3-26}$$

$$z_{km}^{p} \leq x_{mm} \quad \forall k,m \in H\,,\ k \neq m\,,\ p \in \{1,2\ldots,P\} \tag{3-27}$$

$$z_{km}^{p} \in \{0,1\} \quad \forall k,m \in H\,,\ p \in \{1,2\ldots,P\} \tag{3-28}$$

$$z_{kk}^{p} = 0 \ \forall k \in H \ \ p \in \{1,2\ldots,P\} \tag{3-29}$$

Also Constraints (3-2)-(3-9), (3-17)-(3-20).

Constraints (3-22) calculate the flow on the inter-hub link *(k, m)*. Constraints (3-23) and (3.24) guarantee the flow falls in the feasible domain of the corresponding air service. Constraints (3-25) force the model to choose only one service for each hub link. Constraints (3-26) and (3-27) guarantee that this concave cost function is applied only to inter-hub links. Constraints (3-28) force the decision variables to be binary. Constraints (3-29) denote that when the origin hub and the destination hub are the same, air service is not necessary. Constraints (3-2)-(3-9) have the same meaning as those in the basic model. After the flow on the inter-hub link (*k*, *m*) is calculated, the only one cost function that charges the least cost is identified by minimizing the objective function. The corresponding binary variable z_{km}^{p} is set to be 1, in which the index *p* determines the optimal air service, i.e. the fixed cost (intercept k_p) and its corresponding variable cost (slope χ_p), in the objective function.

HLP with cost select function studied in this dissertation shares some similarities with previous researches. The most similar previous research is the FLOWLOC model by O'Kelly and Bryan[221]. Our model differs from the FLOWLOC model in several major aspects.

(1) The biggest difference between the FLOWLOC model and ours is that FLOWLOC is a multi-allocation one, while ours is a single-allocation one;

The single-allocation criterion and flow-dependent cost functions are especially appropriate for air freight networks, since the opportunity to maximize load factors regardless of routing gives the carrier every incentive to capture EOS[222]. When the inter-hub cost function is concave and the single-allocation criterion is applied, the total network cost is minimized by forcing some interacting pairs to use non–least-cost paths. Passenger inconvenience, e.g. longer travel time and congestion at airports, makes such network inappropriate for passenger airlines[223]. As far as solution method is concerned, even when the hub location is fixed, the single-allocation sub-problem remains *NP*-hard[224], while the multi-allocation sub-problem can be converted to an uncapacitated facility location problem (UFLP)[225].

(2) The piecewise linear cost function in the FLOWLOC model is an approximation of a nonlinear function, while each piece of the cost function in this dissertation denotes one kind of air service. All the intercepts, slopes and turning points have economic meaning.

---

[221] See O'Kelly/ Bryan (1998), pp.605-616.

[222] See O'Kelly (1998a), pp. 171-186.

[223] See O'Kelly (1998), p.610.

[224] It has been proved by Kara (1999).

[225] See Klincewicz (2002), pp.107-122.

Another similar research is defined as hub arc location problems by Campbell et al.[226]. The authors studied models that minimize the total cost by selecting so-called hub arcs, which connect two hubs and on which cost rate is discounted by α . In other words, whether the cost rate between two demand nodes is discounted by α depends on the model. The total number of such discounted hub arcs is predefined in the model. The most significant difference between hub arc location problems and ours (also most HLPs) is that hub arc location problem takes an arc-oriented rather than node-oriented point of view and locates certain number of discounted arcs rather than locates certain number of hub nodes. As we can see, in hub arc location problem the cost rate is dependent on arc type but independent on the flow through the arc.

Podnar et al[227] also questioned traditional hub location models and studied modeling approaches, by which the discount rate α is applied under flow-dependent circumstances, i.e. if the flow through a link exceeds the prescribed threshold, the cost is discounted by α . The model also focuses on the links rather nodes. However, both hub arc location problem and model by Podnar et al are still based on the fixed discount factor α , which is not a convincing approach to model EOS. Moreover, such cost structure can only define two cost rates.

Among all the available researches on HLPs with piecewise linear cost function, we have not found any one that points out the irrational gap in the current definition of piecewise linear cost function[228] and makes corresponding remedy. We seem to be the first one to treat this problem through the formulation of a cost select function that can be easily transform into a piecewise linear cost function when all cost parameters are defined.

3.4.2. Extension model 2

As we have mentioned at the beginning of this chapter, the network planning studied in this dissertation is based on the current network. Since Company A has no intention to purchase as many aircraft as necessary for the new network in the near future, the number and the type of current self-owned aircraft should be also considered in the model.

First, we would like to check if the constraints on currently self-owned aircraft fleet can distort the hub location decisions in the long run. Second, we also want to check how these constraints affect the demand allocation decisions. Later, daughter companies will be established according to the decisions on hub location and demand allocation. Each daughter company is responsible for one hub region. For this reason, the allocation decisions are regarded by the management as long-term decisions, whose change will involve a series of issues, such as the change of the shareholding. Third, we would like to provide the self-owned aircraft fleet updating strategy, which is based on the assumption that all the hubs are linked by direct flight.

On accounts of these three motivations, we insert the numerical constraints on the current aircraft in *Ext.1*to formulate extension mode 2 (*Ext.2*). We introduce the parameters listed in Tab. 3-7 to distinguish the air service provided by self-owned aircraft and that from commercial air freight market.

[226] See Campbell et al. (2005a), pp.1540-1555; Campbell et al. (2005b), pp.1556-1571.

[227] See Podnar et al. (2002), pp. 371-386.

[228] See Gap2 in Fig.3-9.

Decisions	Description
S	Set of service by self-owned aircraft $S \subset P$,
C	Set of service from air freight market $C \subset P$, $C \cup S = P$, $C \cap S = \varnothing$
q_s	The number of self-owned aircraft type $s\,(s \in S)$

Table 3-7: Parameters for *Ext.2*

All the assumptions for *Ext.1* are also applied here. Moreover, we have the following assumptions especially for *Ext.2*.

(1) The number of the self-owned aircraft is not enough to fully connect all hubs. Suppose hub number is $h = \sum_{k \in H} x_{kk}$, then $\sum_{s \in S} q_s < (h-1) \times h / 2$.

(2) All self-owned aircraft can be used in the new network. In other words, the situation that some aircraft cannot be used in the new network due to the capacity constraint is eliminated.

(3) Commercial air freight service is available on all inter-hub links. In other words, there are no constraints on q_c .

The *Ext.2* is formulated as follows.

Minimize (3-21)

S.T. $\sum_{k \in H} \sum_{m \in H} z_{km}^{s} = q_s \quad s \in S \subset \{1, 2, ..., P\}$ (3-30)

Also Constraints (3-17)-(3-20), (3-23)- (3-29).

Constraints (3-30) impose numerical constraints on each type of self-owned aircraft.

3.5. Summary

This chapter is dedicated to the formulation of the basic model and its extensions.

In Sec.3.1 we propose a fully interconnected/star shaped H/S network for the multi-modal, time-definite nationwide trans-city overnight EDS. Particularly, dozens of cities in the potential hub set are chosen as hubs, while all the other cities belonging to the target market are allocated to one of these hubs subject to the constraints of maximum hub coverage radius. All hubs also serve as gateways for the air network, which are fully connected by direct flight. Non-hub cities are connected to their "home" hubs by direct ground service. Based on the network, we illustrate 8 different parcel paths and specify the services in the target market.

In Sec.3.2 we propose the basic model, which is formulated as a 0-1 integer program with the four script formulation method.

In Sec.3.3 we base the planning work on the existing network by considering current facilities with "Sunk Cost Theory" and current self-owned aircraft fleet with flow-dependent cost function. To distinguish different air services, a cost select function is adopted, which can be easily transformed into a piecewise linear cost function when all cost parameters are defined. We point out the irrationality in current piecewise linear cost function that is widely applied in previous studies and make a special smooth treatment to remedy this problem.

In Sec. 3.4 we propose two extension models by modifying the air freight cost with the cost selection function proposed in Sec.3.3.3. *Ext.1* determines the optimal air freight on each hub link besides the hub location and demand allocation decisions. *Ext.2* further considers the constraints on current self-owned aircraft fleet based on the assumption that all hubs are connected by direct flight. It aims to check whether these constraints can affect the hub location and allocation decisions and provide aircraft fleet updating strategy. We also point out the difference between our model and other HLPs with flow-dependent cost function in previous studies.

4. Solution design and improvement techniques

4.1. Solution review and design

4.1.1. Literature review on solutions of related problems

As we have mentioned in Sec.2.3.2, up till now only two published studies involve similar models as ours- one is named as Hub Covering Flow Problem (HCFP) by Sim and the other one is by Campbell. Sim solved the 50-node instances from the Australia Post (AP) data set by using the commercial solver Xpress-MP and 25-node instances from the CAB data set by CPLEX.[229] The tests by Campbell are based on the CAB data set with CPLEX.[230] However, it is beyond the capability of CPLEX to solve the large-scale instances with 281 demand nodes from the project.[231]. In the following, we make reviews on solutions for related problems to enlighten our research.

Since the models in this dissertation are combinations of hub location problems with fixed cost and hub set covering problems, we try to list all studies we can find on these two problems in Tab.4-1 and 4-2. We have also referred several recent literature reviews, including those by Alumur & Kara and Hekmatfar & Pishvaee[232]. We briefly describe models and solution techniques. Hub location problems with fixed cost are categorized according to the allocation criterion and hub capacity constraints. Hub capacity constraints are exclusive for the ramification of HLPs, since the number of hubs is not fixed. Hub capacity is considered in former studies in two different ways. One is constraint on the traffic that passes through hubs[233]. The other is constraint on the number of demand nodes that can be connected to hubs.

Problem category	Author and reference	Model and solution technique
Single allocation/ uncapacitated	O'Kelly [234]	quadratic integer program, heuristic algorithm to estimate a good upper bound
	Campbell [235]	the first linear programming formulation
	Abdinnour-Helm /Venkataramanan [236]	new quadratic formulation, branch and bound (B&B) procedure to obtain lower bounds, GAs

[229] See Sim (2007), available on internet: http://ir.uiowa.edu/etd/124. (access on 20.01.2013).

[230] See Campbell (2009), pp. 3107-3116.

[231] IBM ILOG CPLEX Optimizer can solve problems with millions of constraints and variables. However, the numbers of variables and constraints are billions under the instances with 281 demand nodes. See the official website of IBM ilog cplex: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/ (access on 20.02.2013).

[232] See Alumur/Kara (2008), pp.1-21; Hekmatfar/ Pishvaee (2009), pp.243-270.

[233] See e.g. Yaman / Carello (2005), p.3230.

[234] See O'Kelly (1992), pp.292–306.

[235] See Campbell (1994b), pp. 387–405.

	Abdinnour-Helm [237]	hybrid heuristic algorithm of GAs and tabu search (TS)
	Labbe /Yaman[238]	valid and facet defining inequalities
	Topcouglu et al[239]	GAs
	Chen [240]	TS and SA method
	Cunha/ Silva [241]	hub discount rate according to traffic volume, GAs combined with simulated annealing (SA)
	Thomadsen/ Larsen[242]	Branch-and-Price (combination of column generation and set partitioning)
	Silva/ Cunha[243]	Multi-start TS and two-stage integrated TS
Single allocation/capacitated	Campbell[244]	first linear integer formulation
	Ernst /Krishnamoorthy [245]	new formulation, two heuristics, B&B algorithm
	Labbe et al[246]	B&B algortithm
	Costa et al [247]	bi-criteria problems minimizing total cost and service time, interactive decision-aid approach developed for bi-criteria integer linear programming problems to generate non-dominated solutions
Multi-allocation / uncapacitated	Campbell[248]	first linear integer formulation
	Klincewicz[249]	dual ascent and dual adjustment technique within a B&B scheme
	Marianov /Serra[250]	TS
	Mayer /Wagner[251]	new B&B method, dual ascent approach
	Boland et al [252]	preprocess procedures, tight constraints
	Hamacher et al[253]	polyhedral study, new formulation

[236] See Abdinnour-Helm/ Venkataramanan (1998), pp.31–50.

[237] See Abdinnour-Helm (1998), pp. 489–499.

[238] See Labbe et al (2005), pp. 371–405.

[239] See Topcuoglu et al (2005), pp.467–984.

[240] See Chen (2007), pp.211–220.

[241] See Cunha /Silva (2007), pp.747–758.

[242] See Thomadsen/ Larsen (2007), pp.2520–2531.

[243] See Silva/ Cunha (2009), pp.3152–3165.

[244] See Campbell (1994b), pp.387–405.

[245] See Ernst/ Krishnamoorthy (1999), pp. 141–159.

[246] See Labbe et al (2005), pp. 371–405.

[247] See Costa et al (2008), pp.3671–3695.

[248] See Campbell (1994b), pp.387-405.

[249] See Klincewicz (1996), pp.173–184.

[250] See Marianov/ Serra (2003), pp.983–1003.

[251] See Mayer/ Wagner (2002), pp.715–739.

[252] See Boland et al (2004), pp.638–653.

	Marin [254]	valid inequalities, relax-and-cut algorithm
	Marin et al [255]	new universal formulation, preprocess to reduce problem size
	Canovas et al[256]	heuristic-based and dual ascent technique, B&B algorithm
	Camargo et al[257]	benders decomposition
Multiple allocation/ capacitated	Campbell[258]	first linear integer formulation
	Ebery et al[259]	new formulation, customized heuristic, B& B algorithm
	Boland et al[260]	reprocess procedures, tight constraints
	Marin [261]	tight integer linear programming formulations with some properties of the optimal solutions to speed up the solution
	Rodriguez-Martin/ Salazar-Gonzalez[262]	mixed integer linear programming with branch-and-cut based on decomposition method

Table 4-1: Literature review on hub location problems with fixed cost[263]

The studies on HLPs with endogenous hub number, mainly hub location problems with fixed cost and hub set covering problems, are not so rich compared to HLPs with exogenous hub number p, mainly p-hub median problem and p-hub center problem[264]. Nevertheless, it is also indicated clearly by the review that there is a significant increase of such researches since the year 2004, most of which resort to meta-heuristics[265].

Moreover, compared to models with multi-allocation criterion, models with single-allocation criterion is more difficult to solve. In the case of problems with multi-allocation and no capacity constraints, once the location of hubs is fixed, the allocation problem can be efficiently solved with an all-pairs shortest path/least cost problem[266]. However, in the corresponding problem with single-allocation criterion, even when hub location is

253 See Hamacher et al (2004), pp.104–116.

254 See Marin (2005), pp.393-422.

255 See Marin et al (2006), pp.274–292.

256 See Canovas et al (2007), pp. 990–1007.

257 See De Camargo et al (2008), pp.1047–1064.

258 See Campbell (1994b), pp.387-405.

259 See Ebery et al (2000), pp.614–631.

260 See Boland et al (2004), pp. 638–653.

261 See Marin (2005a), pp.3093–3109.

262 See Rodriguez-Martin/ Salazar-Gonzalez (2008), pp. 468–479.

263 It is to be noted that in the literature review by Hekmatfar & Pishvaee the hub location problem with fixed cost is categorized under the name uncapacitated hub location problem (UHLP) and capacitated hub location problem(CHLP), without mentioning the name hub location problem with fixed cost. It is also to be noted that hub location problems with fixed cost and exogenous hub number p are not included here. See two examples, Aykin (1995), pp.201–221; Sasaki/ Fukushima (2003), pp.409-428. Actually they are the extensions of p-hub median problems by including hub fixed cost in objective function.

264 See Hekmatfar/ Pishvaee (2009), p.260.

265 See Hekmatfar/ Pishvaee (2009), p.261.

266 See Ernst/ Krishnamoorthy (1998), pp.146-162.

known, the resulting allocation problem is equivalent to a quadratic semi-assignment problem, which is proved to be *NP*-hard[267].

Author and Reference	Model and Solution Techniques
Campbell[268]	first mixed integer formulation for hub covering problems with single and multiple allocation
Kara/ Tansel [269]	new formulation for single allocation and multiple allocation with quantity-independent discount rates, three different linearization of quadratic model
Wagner[270]	improved model formulation from Kara and Tansel, reduce problem size by removing redundant constraints
Ernst et al [271]	a novel formulation focused on hub radius, enumerative solution method
Hamacher/ Meyer[272]	identified facet defining valid inequalities, binary search algorithm
Tan/ Kara [273]	latest arrival hub covering formulation, iterative location-routing heuristic
Weng /Wang[274]	improve multiple allocation hub covering problem, scatter search and GAs
Qu/ Weng [275]	multiple allocation hub maximal covering problem, path relinking approach
Hatice Calik et al [276]	single allocation over incomplete hub network, TS
Alumur et al[277]	linearization, incorporate valid inequalities, optimization with CPLEX

Table 4-2: Literature review on hub set covering problems

Tab.4-2 shows that the number of papers on hub set covering problems is even fewer compared with hub location problems with fixed cost reviewed in Tab.4-1. After this problem was proposed by Campbell[278] in 1994, it remained untouched until the year 2003 by Kara and Tansel. They also proved that single-allocation hub set covering problem is *NP*-hard[279].

[267] See Kara (1999). Sohn and Park also proved it NP-hard when the hub number is larger than 2. See Sohn/ Park (2000), pp.17-25.

[268] See Campbell (1994b), pp.387-405.

[269] See Kara/ Tansel (2003), pp.59–64.

[270] See Wagner (2007), pp.932-938.

[271] See Ernst (2005), pp.1-18.

[272] See Hamacher/ Meyer (2006), pp.1-18.

[273] See Tan/ Kara (2007), pp. 28–39.

[274] See Weng/ Wang (2008), pp.408–411.

[275] See Qu/Weng (2009), pp.1890-1894.

[276] See Calik et al (2009), pp. 3088-3096.

[277] See Alumur et al (2009), pp.936–951.

[278] See Campbell (1994b), pp.387-405.

[279] See Kara/ Tansel (2003), pp.59 - 64.

Two categories of solution methods

Most of the HLPs are studied with mixed-integer programming (MIP) models or 0-1 integer programming (IP) models. As we go through solution methods for HLPs, they can be roughly divided into two categories according to the master algorithms[280], i.e. exact methods and heuristics, mainly meta-heuristics[281].

(1) Exact methods

Advanced exact algorithms that have been used for solving HLPs include:

- cutting-plane method,
- branch and bound (B&B),
- branch and cut,
- branch and price,
- Lagrangian relaxation,
- column generation,
- partitioning method,
- Benders' Decomposition,
- dual-ascent and
- combination of these[282].

HLP was first formulated as a quadratic model by O'Kelly[283]. Linearization was a great advance that allows the use of linear programming methods to find the optimal solution and prove its optimality[284]. Techniques

[280] With the development of algorithms for combinatorial optimization problems, more and more algorithms are hybrids of exact methods and (meta-) heuristics, which will be discussed in Sec.4.1.2. For this reason, it is better to categorise solution methods according to the master algorithm. For example, the method by Ernst and Krishnamoorthy is essentially an exact algorithm incorporated with meta-heuristics. They used upper bound obtained from the SA heuristic to develop a LP-based B&B algorithm. See Ernst /Krishnamoorthy (1996), pp. 139–154. Readers who are interested in hybridized algorithms may refer to Raidl/ Puchinger and Jourdan et al. See Raidl/ Puchinger (2008), pp.31-62; Jourdan et al (2009), pp.620-629.

[281] Heuristics can be categorized mainly into three classes: constructive heuristics, improvement heuristics and incomplete exact heuristics. Constructive heuristics generate solutions from scratch by adding opportunely defined solution components to an initially empty partial solution. This is done until a solution is complete or other stopping criteria are satisfied. Improvement heuristics, e.g. local search, start from some initial solution and iteratively try to replace the current solution by a better solution in an appropriately defined neighborhood of the current solution. Incomplete exact heuristics generate feasible solutions for early stage of exact solutions, e.g. B&B. Meta-heuristics are solution methods that orchestrate an interaction between the basic heuristics and higher level strategies to create a process capable of escaping from local optimal and performing a robust search of a solution space. See e.eg Mayer (2001), p.90; Domschke (1997),Chapter1.3; Domschke/ Drexl (1998), p.120; Dorigo / Stützle. http://www.metaheuristics.net/(access on 20.01.2013).

[282] Exact methods can also be combined with other exact methods. For example, Thomadsen and Larsen inserted column generation into branch-and-price algorithm to solve a hierarchical network problem. Canovas et al embeded dual-ascent technique in B&B framework to solve uncapacitated multi-allocation HLP. Rodriguez et al. based branch-and-cut algorithms on Benders' Decomposition method. Sasaki and Fukushima presented a model for the capacitated 1-stop multi-allocation HLP, which was solved by a B&B algorithm with Lagrangean relaxation bounding strategy. See Thomadsen/Larsen J (2007), pp.2520–2531; Canovas et al (2007), pp.990–1007; Rodriguez et al (2007), pp. 495–505; Sasaki/ Fukushima (2003), pp.409-428.

have been proposed to accelerate the solution speed and improve the performance of algorithms, including tailored heuristics in linear programming[285], tightening constraints to improve the linear programming relaxation[286], preprocessing techniques[287], eliminating redundant and impractical routes[288], multi-start nodes for B&B[289], hublocater to obtain lower bounds[290] and analyzing feasibility polyhedron and identifying facet-defining valid inequalities[291].

Solution time with exact methods, however, often increases dramatically with the instance scale. For this reason, exact methods for HLPs are popular for instances with less than 50 nodes. Up till now, we have only found few studies involving medium or large scale hub location problems with fixed cost (i.e. more than 50 nodes) with exact methods, e.g. Wagner (500 nodes)[292], Contreras et al (200 nodes)[293], Ebery et al (200 nodes)[294], Camargo et al (200 nodes)[295] and Canovas et al (120 nodes)[296]. The largest hub covering problem that has been studied with exact methods so far, to the best of our knowledge, is that by Hamacher and Meyer with 50 nodes[297].

(2) (Meta-) heuristics

While exact methods can hardly be employed for large-scale problems without much effort on customized improvement on both models and algorithms, meta-heuristics follow relatively standard solution frameworks

[283] See O'Kelly (1992), pp.292–306.

[284] For example, Campbell proposed the first linear programming formulation. However, Campbell's model resulted in fractional solutions. Skin-Kapov et al (1996) obtained a tight linearized version of the HLP that resulted in integer solutions for the hub locations. Ernst and Krishnamoorthy devised a linearized variation of O'Kelly's quadratic model. See Campbell (1994b), pp. 387-405; Skorin-Kapov et al (1996), pp. 582-593; Ernst/ Krishnamoorthy (1996), pp.139-154.

[285] See Pirkul/ Schilling (1998), pp.235-242; Campbell (1996), pp. 923-935; Ernst/ Krishnamoorthy (1998a), pp.100-112.

[286] See Boland et al. (2004), pp. pp. 638–653.

[287] See Wagner (2007), pp. 932-938.

[288] By eliminating redundant and impractical routes and by exploiting the symmetry of the available test data, O'Kelly et al. modified the model by Skin-Kapov et al, reducing computation time and the number of variables without sacrificing integrality. See O'Kelly et al (1996), pp. 125–138; Skorin-Kapov et al (1996), pp. 582-593.

[289] See Ernst/ Krishnamoorthy (1998b), pp.149-162.

[290] See Mayer/ Wagner (2002), pp.715-739.

[291] See Labbe/ Yaman (2004), pp.84-93; Hamacher et al (2004), pp.104-116; Hamacher/ Meyer (2006), pp.1-18.

[292] See Wagner (2007), pp.391-401.

[293] This paper presents a branch-and-price algorithm for the capacitated single allocation HLP, in which Lagrangean relaxation is used to obtain tight lower bounds of the restricted master problem. This method can solve instances of up to 200 nodes to optimality, which seems to be the largest instances that have been solved for this problem. See Contreras et al (2011), p.41-55.

[294] Ebery et al considered capacitated multi-allocation HLP. The authors solved the problem with 200 nodes by incorporation the upper bound obtained from shortest paths in a linear programming-based B&B solution procedure. See Ebery et al (2000), pp. 614–631.

[295] See Camargo et al (2008), pp. 1047–1064.

[296] The paper deals with the uncapacitated multi-allocation HLP. The authors designed a heuristic method based on a dual-ascent technique to embed in an exact B&B framework. This algorithm can solve instances with up to 120 nodes. See Canovas et al (2007), pp. 990-1007.

[297] See Hamacher/ Meyer (2006), pp.1-18.

and can provide good solutions within reasonable time by sacrificing the guarantee of finding optimal solutions.

Meta-heuristics are solution methods that orchestrate an interaction between the basic heuristics and higher level strategies to create a process capable of escaping from local optimal and performing a robust search of a solution space[298]. The class of meta-heuristics includes-but is not restricted to- Simulated annealing (SA), tabu search (TS), genetic algorithms(GAs), ant colony optimization, scatter search, path relinking, greedy randomized adaptive search procedure (GRASP), multi-start methods, guided local search (GLS) and variable neighborhood search(VNS).

Generally speaking, meta-heuristics can handle much larger problems than exact methods, although no optimal solutions are guaranteed. The computational time always increases with the instance scale mildly with meta-heuristics rather than exponentially with exact methods. Second, meta-heuristics are more flexible in the sense that additional concerns can be incorporated into problems without much increase in the running time and deterioration to the solution quality. Third, meta-heuristics provide uniform solution framework so that different problems can be easily shaped into the framework by only changing parameters accordingly. Without comprehensive understanding of the specific model, one can still obtain relatively good solutions in a reasonable time. For these reasons, meta-heuristics have been regarded as an ideal instrument to solve large and complicated *NP*-hard problems. Actually, research on large-scale combinatorial optimization problems was not abundant and intensive until recently, after fast development of meta-heuristics[299].

Heuristics and meta-heuristics for HLPs

HLP, to be precise, single-allocation p-hub median problem, was first formulated as a quadratic integer program by O'Kelly. The author was also the first one to develop two heuristics to solve it[300]. Klincewicz [301] solved this problem by using clustering heuristics and TS. Later, further improvement on location decisions was made with TS and GRASP[302]. Skorin-Kapov and Skorin-Kapov [303] proposed a TS heuristic for both location and allocation phases. Aykin[304] developed a SA-based interchange heuristic to solve this problem. Ernst and Krishnamoorthy[305] also resorted to SA heuristic. Smith et al [306] used neural networks, which yielded disappointing results.

[298] See Gendreau/ Potvin (2010), p.vii. Other definitions proposed in the literature please see Dorigo/ Stützle (2000); Osman/ Laporte (1996), pp.513-623; Voß et al (1999); Stützle (1999). For detailed introduction to meta-heuristics, please refer to Blum/ Roli (2003), pp.268–308.

[299] See Hertz/Widmer (2003), p.247.

[300] One is to assign all non-hub nodes to their nearest hub, whereas the other is to assign to either nearest or second-nearest hub. See O'Kelly (1987), pp.393-404.

[301] See Klincewicz (1991), pp.25-37.

[302] See Klincewicz (1992), pp.283–302.

[303] See Skorin-Kapov/ Skorin-Kapov (1994), pp. 502-509.

[304] See Aykin (1995b), pp.200-219.

[305] See Ernst/ Krishnamoorthy (1996), pp. 139-154.

[306] See Smith et al (1996), pp.155-171.

Since (meta-) heuristics can neither guarantee optimal solutions nor prove their optimality, researches on the solution of HLPs shifted to linearization, new formulations and linear programming algorithms to find optimal solutions and to prove the efficiency of meta-heuristics by offering benchmark. Recently, especially after 2005 the research focus seems to come back to meta-heuristics perhaps out of the requirement to solve large-scale and complicated real-life HLPs[307]. More and more hybrid ones came into being, taking advantage of the strengths of each individual meta-heuristic components to explore the solution space better. Hybrids of meta-heuristics with other exact optimization algorithms, like B&B, are also increasingly popular[308].

In the following, we make a brief review on the meta-heuristics for HLPs with endogenous hub number- hub location problem with fixed cost and hub set covering problem.

(1) Hub location problem with fixed cost

Abdinnour-Helm proposed a heuristic method for UHLP-S based on a hybrid of GAs and TS.[309] Firstly, GAs is used to determine the number and the location of hubs and then each demand point is assigned to its closest hub to generate an initial solution for the TS heuristic which finds the optimal allocation decisions. The author compared her results with the GAs by Abdinnour-Helm and Venkataramanan[310] and found that the algorithms using TS in combination with GAs performance much better algorithms using GAs alone.

Topcuoglu et al. proposed a GAs-based method for the UHLP-S.[311] Each chromosome in GAs consists of two arrays: HubArray and AssignArray. The lengths of these arrays are equal to the number of nodes. They applied roulette sampling for fitness selection, single-point crossover operator and two mutation operators- shift and exchange. The authors compared their solutions with the best solutions presented in the literature and demonstrated that both the solution quality and the running time surpass former works with CAB data set (small-scale problem) and AP data set (large-scale problem).

Cunha and Silva[312] combined GAs with SA. Specifically, after GAs determines the hub location, the demand points are first allocated to the nearest hub. LS with shift and swap movements is then applied for the allocation decisions. SA is also incorporated into LS to prevent it from being stuck at local optimal. This hybrid heuristic was proved to outperform the GAs by Abdinnour-Helm[313].

Another algorithm for this problem was proposed by Chen[314]. His hybrid heuristic is based on the SA method, tabu list and improvement procedures, named by the author SATLUHLP. The proposed heuristics are divided into three levels: the first level is to determine the number of hubs; the second level is to select hub locations for a given number of hubs; the third level is to allocate the non-hubs to the chosen hubs. Specifically, the up-

---

[307] See Cunha /Silva (2007), pp.747–758; Chen (2007), pp.211–220; Silva/ Cunha (2009), pp.3152–3165.

[308] See e.g. two recent reviews on this topic: Raidl/ Puchinger (2008), pp.31-62; Jourdan et al (2009), pp.620-629.

[309] See Abdinnour-Helm (1998), pp. 489–499.

[310] See Abdinnour-Helm/ Venkataramanan (1998), pp. 31-50.

[311] See Topcuoglu et al (2005), pp. 967-984.

[312] See Cunha /Silva (2007), pp. 747-758.

[313] See Abdinnour-Helm (1998), pp. 489–499.

[314] See Chen (2007), pp.211-220.

per bound for the number of hubs is determined with SA mechanism by calculating the trend of the marginal reduction of the transportation cost. Restricted single location exchange procedure with tabu list is applied for hub location level. At the third level, non-hub nodes are allocated to the nearest hub followed by a LS procedure. The tests under medium and large scale instances (100-node and 200-node instances with the AP data set) showed that SATLUHLP could yield the best known solutions with less running time.

The latest research on this problem was conducted by Silva and Cunha[315]. Three variants of a multi-start TS heuristics as well as a two-stage integrated TS heuristic were proposed. With multi-start heuristics several different initial solutions are constructed and then improved by TS, while the two-stage integrated TS heuristic is applied to improve both the location and allocation part of the problem. Computational tests with the CAB data set and the AP data set showed that these approaches consistently returned the best-known results in very short running time. The authors also reported the integer optimal solutions for all 80 CAB data set instances and the 12 AP data set instances up to 100 nodes

(2) Hub set covering problem

To the best of our knowledge, up till now there are only three studies on hub covering problem using meta-heuristics. Meanwhile, one of them is maximal hub covering problem with exogenous hub number[316]. We list the other two hub set covering problems in the following.

Weng and Wang [317] considered the multi-allocation hub set covering problem. The study provided two evolutionary approaches by scatter search and GAs. The computational tests show that GAs get a better performance than scatter search with perspective of both solution quality and computational time.

Calik et al[318] considered the HLRP for postal delivery systems and developed an iterative two-stage solution procedure. In the first stage, hub locations are determined and postal offices are multiply allocated to the hubs. In the second stage, routes in hub regions are planned to alter the distances used in the hub location problem. The procedure then iterates between the two stages by updating the distances used in the hub location problem till certain termination criterion is satisfied.

On the one hand, the tremendous number of variables and constraints under large-scale instances restricts the use of exact method only to small instances. On the other hand, meta-heuristics can yield a few near-optimal solutions for large-scale instance in a reasonable time. Moreover, they have standard overall procedures that are not only easy to use but also flexible to be tailored with problem-specific knowledge.

Meta-heuristics will be adopted in this dissertation under several specific considerations.

---

[315] See Silva/ Cunha (2009), pp.3152–3165.

[316] Qu and Weng considered the multi-allocation hub maximal covering problem. The authors provided an evolutionary approach based on path relinking. The Computational experiences were based an AP data and on hub airports location of Chinese aerial freight flows between 82 cities. See Qu/ Weng (2009), pp.1890-1894.

[317] See Weng/ Wang (2008), pp.408–411.

[318] See Calik et al (2009), pp. 3088-3096.

(1) Instance scale

The overwhelming reason for us to adopt meta-heuristics is the scale of our real-life instances. The aim of our research is to design a multi-modal H/S network for nationwide trans-city overnight EDS in China. The problem is to select hubs from 281demand nodes and singly allocate the rest demand nodes to hubs, belonging to the few largest instances among our available studies. Almost all the tests in former studies on HLPs under large-scale instances are based on the AP data set (maximum 200 nodes) with few exceptions[319]. Under the consideration of instance scale, meta-heuristics is an ideal instrument for us to obtain near optimal solutions within reasonable computational time.

(2) Management preference

In order to support strategic decision, our research is much better to offer managers several near-optimal solutions for tradeoff than the only one optimal solution with the minimum total cost of the network we can find. Strategic decision-makers must consider not only the cost but also other factors, such as organization structure, local economic situation and governmental policies. With meta-heuristics the management can thus choose from several near-optimal solutions with different hub location and demand allocation decisions by considering other managerial and social factors. These solutions, in perspective of managers, may be more attractive, although they result in higher cost than the best one.

4.1.2. Hybrids of meta-heuristics

For combinatorial optimization problems (COPs) that are *NP*-hard, no polynomial time algorithm exists, assuming that $P \neq NP$. Therefore, complete methods might need exponential computational time in the worst-case.[320] By using approximate methods, such as meta-heuristics, we sacrifice the guarantee of finding optimal solutions for the sake of getting good solutions in a significantly reduced amount of time. Thus, the use of meta-heuristics has received more and more attention in the last 30 years. In the first two decades the applications were confined to rather standard meta-heuristics. However, recent researches have shown that a skilled combination of a meta-heuristic with other optimization techniques, a so-called hybrid meta-heuristic, can provide a more efficient behavior and a higher flexibility when dealing with real-world and large-scale problem[321].

- **Classification of hybrid meta-heuristics**

Nowadays we can observe a common agreement on the advantage of combining components from different search techniques. The tendency of designing hybrid techniques is widespread in the fields of OR and artificial intelligence (AI), mostly based on the no free lunch theorems[322].

[319] For example, Wagner's exact solution procedure for HLPs with 500 nodes, and Resende & Werneck solved the uncapacitated facility location problem with hybrid multi-start heuristic under instance with 1000 nodes. See Wagner (2007), pp.391-401;.Resende/ Werneck (2006), pp.54-68.

[320] See Blum/Roli (2008), p.1.

[321] See Blum/Roli (2008), p.1.

[322] See Wolpert. Marready (1997), pp.67-82; Raidal (2006), p.3.

We may distinguish hybrid heuristics in two dimensions, i.e. hybrid contents and hybrid level. A matrix is adopted to illustrate their relationships (see Fig. 4-1).

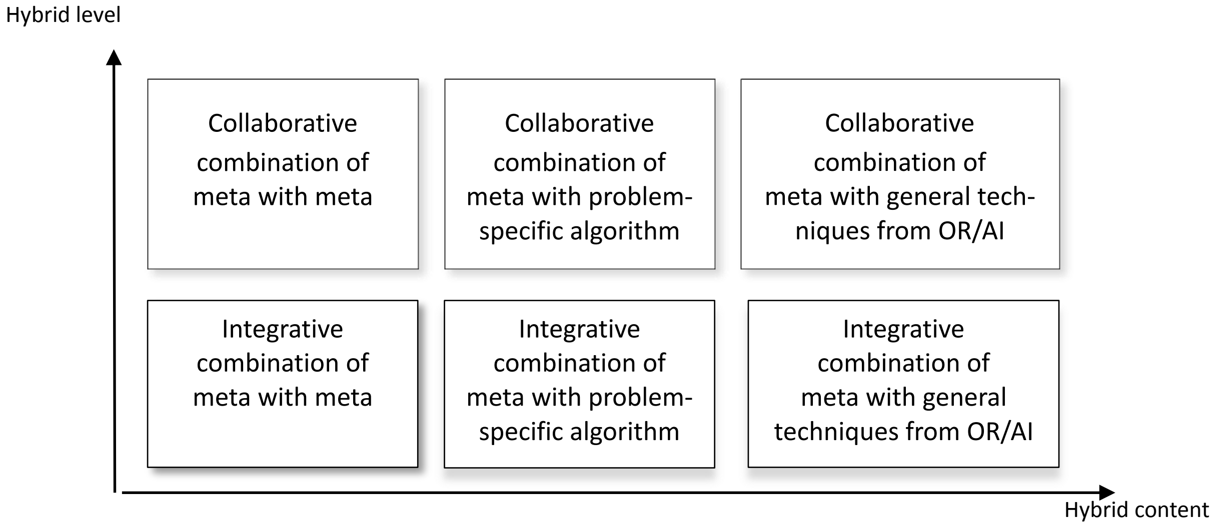

Figure 4-1: Classification of hybrid meta-heuristics

Dimension of hybrid contents

In terms of hybrid contents, we may distinguish between three categories: the first one combines meta-heuristics strategies; the second one combines meta-heuristics with certain algorithms specific for the problem; the third one combines meta-heuristics with other more general techniques coming from fields like OR and AI.[323]

A prominent example of the first category is the use of trajectory methods[324] into population-based techniques[325]. The reason becomes apparent by analyzing the respective strengths of trajectory methods and population-based methods. The power of population-based methods lies in the capability of recombining solutions to obtain new ones. This enables the search process to perform a guided sampling of the search space and identify promising areas. Meanwhile, the strength of trajectory methods lies in the way they explore a promising region in the search space. A promising area in the search space is searched by trajectory methods in a more structured way than by population-based methods so that the search is driven towards local optima or confined areas of the space in which many local optima are condensed[326]. In sum, population-based methods are better in identifying promising areas in the search space, from which trajectory methods can quickly reach good local optima. Therefore, meta-heuristic hybrids can effectively combine the strengths of both population-

[323] See Raidl (2006), p.4.

[324] Generally speaking, algorithms that work on a single solution are referred to as trajectory methods. They comprise all meta-heuristics that are based on local search, such as TS, iterated local search and variable neighborhood search. See Blum/Roli (2008), p.6.

[325] Population-based meta-heuristics deal at each algorithm iteration with a set of solutions rather than with a single solution. From this set of solutions the population of the next iteration is produced by the application of certain operators.

[326] See Chiarandini et al (2006), p. 118.

based methods and trajectory methods. Successful examples for the third category are hybrids of meta-heuristics with OR methods, such as linear programming[327], branch & bound, tree-based search techniques[328], dynamic programming and neutral networks. A recent literature review on this topic was made by Jourdan et al[329].

Dimension of hybrid level

In terms of hybrid level, we may distinguish between two categories: collaborative combinations and integrative combinations[330].

Collaborative combinations are based on the exchange of information about states, models, entire sub-problems, solutions or search space characteristics between several optimization techniques run sequentially (or in parallel). This kind of combination is more related to cooperative and parallel search and it in principle retain the individual identities of the original algorithms[331]. On the contrary, original algorithms in integrative combinations strongly depend on each other. One technique is a subordinate or embedded component of the other technique. Thus, there is a distinguished master algorithm, and at least one integrated algorithms[332].

- **Hybrid principle**

The motivation of hybridization of different algorithmic concepts is to obtain systems with better performance by exploiting and uniting advantages of the individual algorithm[333]. Hybridization is also a way to inject problem-specific knowledge according to No-Free-Lunch Theorem by Wolpert and Macready[334]. Of great importance of hybridization is the dynamic balance between intensification and diversification, i.e. local exploitation and global exploration. They are two contrary but also complementary forces that largely determine the effectiveness of the algorithms[335].In other words, local and intensive exploitation focuses on examining neighbors of elite solutions, while global and extensive exploration is to encourage the search process to examine unvisited regions and to generate different solutions. Therefore, the hybrid meta-heuristics can, on the one side, quickly identify regions in search space with high quality solutions and, on the other side, not waste too much time in regions of search space which have already been explored or which do not provide high quality solutions.

---

[327] Linear programming is often used either to solve a sub-problem or to provide dual information to a meta-heuristic in order to select the most promising candidate solution or solution component. See e.g. Blum (2005), pp.1565-1591; Ibaraki/ Nakamura (2006), pp.13-27; Maniezzo (1999), pp.358-369.

[328] See e.g. Focacci et al (2003), pp.369-403.

[329] See Jourdan et al (2009), pp.620-629.

[330] See e.g. Raidl (2006), p.3; Puchinger/ Raidl (2005), pp.41-53; Jourdan et al (2009), pp.620-629. Collaborative combination is also called parallel, cooperative or high level combination. Integrative combination is also called low level combination.

[331] For interested readers, please refer to Alba (2005); Grainic/ Toulouse (2002), pp.247-249; Sondergeld/ Voß (1999), pp.297-312.

[332] See Talbi (2002), p.543; Puchinger/ Raidl (2005), p.42.

[333] See Puchinger/ Raidl (2005), p.42.

[334] See Wolpert / Macready (1997), p.68.

[335] See Yagiura/ Ibaraki (2001), pp.33-55.

4.1.3. Solution process review and design

The overview of meta-heuristics in last section indicates that it is not problem-specific but an iterative master process that guides and modifies the operations of subordinate algorithms to find better solutions[336]. In this section, after we review the solution process of compound location problems, we divide the original problem into several hierarchical sub-problems and propose a framework of overall solution process, in which the meta-heuristics guides the subordinate algorithms that are designed specifically for each sub-problem.

3.3.3.1 Literature review on solution process of compound location problems

Constrained by solution technique or understanding to the problem, earliest researches focused on one single problem, either location, allocation or routing problem, in the network under strict assumptions of other influencing factors. For example, some studies focused on VRPs in tributary network with the assumption that hubs are fully interconnected[337], while others planed the backbone network with a predetermined set of hubs[338]. Later, researchers proposed compound problems by realizing that these problems are actually interrelated.

On accounts of the complexity of compound location problems, it is a common way to divide the whole problem into several sub-problems (or stages), which are easier to solve separately than the former complete one. Different sub-problems of HLP/ FLP take up different positions in the solution process of the complete problem with the evolution of understanding to the involving problems. In the following, we make literature review on this issue. As researches on HLRPs began much later than those on LRPs[339], most of the literatures we cite here are LRPs.

Sequential method

Sequential methods were first introduced. For example, Jacobson/ Madsen and Nambier et al[340] first solved the location problem by minimizing the sum of hub-to-customer distances and then solved the resulting route-planning problem based on the hub location decision. A more sophisticated method is to estimate beforehand the route length connecting hub and customers with certain formulation for location problem[341]. However, since there is indeed no feedback or information exchange between location problem and routing problem, suboptimal solution for the problem is inevitable[342].

Cluster method

[336] See Voss et al (1999), p. i.

[337] For example, Gavish and Balakrishnan et al studied algorithms for tributary network design. See Gavish (1991), pp.17-71; Balakrishnan et al (1991), pp.237-284; Balakrishnan et al (1995), pp.58-76.

[338] See e.g. Agarwal (1989), pp.64-76; Altinkemer/Yu (1992), pp. 365-381; Balakrishnan/ Altinkemer (1992), pp.192-205; Baybars/ Edahl (1988), pp. 503-528; Chang /Gavish (1993), pp.99-131; Gavish /Altinkemer (1990), pp. 236-245.

[339] The earliest research on HLRP we can find is from Nagy and Salhi. See Nagy /Salhi (1998), pp. 261 – 275. However, research on LRP began as early as 1970s. See e.g. Waston-Gandy/ Dohrn (1973), pp. 321-329; Or/ Pierskalla (1979), pp. 86-94; Bednar/ Strohmeier (1979), pp.89-104.

[340] See Jacobson/Madsen (1980), pp. 378–387; Nambier et al. (1989), pp.14–26.

[341] See Daganzo (2005)

[342] See Balakrishnan et al (1987), p.37; Salhi/ Rand (1989), pp.150-156; Sahlhi/ Nagy (1999), pp.3-19.

In contrast to up-down approach as sequential method, cluster method takes up down-up process by grouping nodes into clusters with some statistical techniques. It first partitions the customer set into clusters, then locates a facility /hub in each cluster and finally solves VRP for each cluster[343]. Just as sequential method, no feedback takes place.

Iterative method

Realizing this pitfall, Bookbinder& Reece and Perl& Daskin[344] introduced iterative method. This method iteratively solve location, allocation, backbone routing and tributary routing problems by feeding information from one phase to another until some stopping criteria are met. A typical iterative framework is as follows[345].

Step 1: Choose hub nodes, perhaps based on available information from an incumbent solution

Step 2: Assign non-hub nodes to tributary networks

Step 3: Design the tributary networks

Step 4: Design the backbone network

Step 5: Evaluate the solution

Step 6: If solution is acceptable, stop. Otherwise, feed information of current solution back to step 1, and repeat the process.

This method has been demonstrated to improve the solution quality compared with that of a sequential method[346]. It is widely adopted by many studies with variations and omission[347].

Hierarchical method

Nagy and Salhiargued that hierarchical method arose because the FLP (here also HLP) is essentially a location problem that takes the routing decision into consideration as well, and this leads to a hierarchy between locating and routing[348]. The idea conforms to the view of management that location problem is the crux that has much longer planning horizon than allocation and routing problems. Based on the concept of "nested method", they proposed a hierarchical heuristic solution framework, in which the master algorithm is devoted to solving the location problem and refers in each step to a subordinate heuristic that solves the routing problem. They

[343] See e.g. Kleinrock/ Kamoun (1980), pp. 221-248; Klincewicz (1991), pp.25-37; Saha/ Mukherjee (1995), pp.378-383. Recent researches include e.g. Barreto et al (2007), pp. 968-977; Billionnet et al (2005), pp.968-977.

[344] See Bookbinder/Reece (1988), pp. 204–213; Perl/Daskin (1985), pp. 381–396.

[345] See Klincewicz (1998), p.314.

[346] See e.g. Salhi/ Nagy (2009), pp. 287-296.

[347] See e.g. Boorstyn/ Frank (1977), pp.29-47; Gerla/ Kleinrock (1977), pp. 48-60; Gavish (1982), pp.355-377; Monma /Sheng (1986), pp.946-965; Lin/ Rath (1987), pp. 18-25; Gavish (1992), pp.167-191.

[348] See Nagy/Salhi (1996a), pp.1166–1174; Nagy/Salhi (2007), pp.649-672.

reported a 6% improvement from hierarchical over the sequential approach but with longer computation time[349].

Both iterative and hierarchical methods give feedback between sub-problems. However, sub-problems in iterative method are independent and parallel and are treated with equal importance, while location problem in hierarchical method is usually taken as the master problem which is incorporated with other sub-problems. In our opinion, the fundamental difference between hierarchical method and iterative method is the division approach for the sub-problems. In iterative method one sub-problem may be contrary in certain aspect to another one, while in hierarchical method the sub-problem at lower stage is coherent with those at higher stage. Moreover, the result of the problem at higher stage depends on the outcome from lower stage. Sub-problems in other methods mentioned above can be solved individually, while the sub-problem at higher stage in hierarchical method cannot be solved until the problems at lower stage are solved. We would like to explain this with an example that one HLP consists of two decisions: (1) where to locate the hubs so as to minimize the transportation costs; (2) how to construct a routing system for tributary network so as to satisfy the given service level. The iterative two-stage solution process proposed in the study repeatedly updates the transportation cost rate in the location problem with the result from routing problem. In other words, the location sub-problem with direct spoke assumption is solved repeatedly by modifying the estimation of feeder link cost rate from the multi-stop routing problem. However, when this problem is solved with hierarchical method, the location problem must include routing decision rather than take direct feeder link as assumption. So the location problem cannot be solved until the feeder routes are determined.

Actually before Nagy and Salhi gave it the well-acknowledged name "hierarchical approach", this method had already been adopted in former studies. For example, Nambiar et al.[350] presented a method that uses the result of their simple depot clustering heuristic as the starting point. Then, they consider in turn $p = 1, 2, \ldots m$ depots being open. For each value of p, they reformulate the LRP as a p-median problem with tour lengths as variable costs and solve it with an exact method. Routing is then solved with a savings method. If the cost of the LRP with p depots is more than that with p-1, the procedure is stopped. This can be viewed as a hierarchical method, since the routing costs are explicitly included in the location model. Nagy and Salhi's[351] "nested method" consists of a location algorithm with LS that refers to a routing method when evaluating neighboring solutions. The location algorithm is based on TS and an add/drop/shift neighborhood. After each move, the routing solution is fully evaluated using a multi-depot VRP algorithm. Lin et al.[352] first determine the minimum number of facilities. Then the VRP solution is completely evaluated for all combinations of facilities. Vehicles are allocated to trips by completely evaluating all allocations. If the best routing cost found is more than the setup cost for an additional depot, the algorithm moves on to evaluating all sets of facilities that contain one more depot. In method by Albareda-Sambola et al[353] an initial solution is found via the linear programming relaxation of the model. The master algorithm for location decision follows the TS framework with

[349] See Nagy/Salhi (1996a), pp.1166–1174.

[350] See Nambiar et al. (1981), pp. 183-189.

[351] See Nagy/Salhi (1996a), pp.1166–1174.

[352] See Lin et al (2002), pp. 5-25.

[353] See Albareda-Sambola et al (2005), pp. 407-428.

LS moves of add, drop and shift. However, infeasible routing solutions are allowed and a penalty is included in the locational objective function for the violation.

3.3.3.2 Solution process design

- **Division of the original problem**

Three decisions are involved in the original problem, i.e. the hub location decision, the allocation decision and the air service selection decision. Actually our original problem is essentially a hub location problem, which is embedded with an allocation problem and an air service selection problem. In the first place, we would like to divide the original problem into three hierarchical sub-problems and clarify the relationship between the three sub-problems and the three decisions in original problem (see Fig. 4-2)

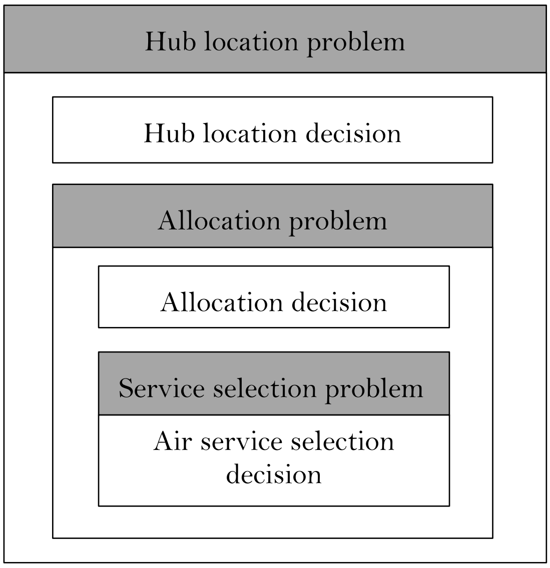

Figure 4-2: Hierarchical sub-problems of the original problem

(1) Hub location problem (upper problem)

The hub location problem takes the objective of minimizing the total cost under the constraint that the distance between all demand nodes and their nearest hub is within the hub maximum coverage. The input of hub location problem includes the location of all demand nodes, potential hub set, demand volume between all O-D pairs, hub fixed cost, transportation cost rate for both backbone and tributary network, and number and capacity of aircraft. In order to calculate the total cost of the network, the hub location problem involves all the three decisions, i.e. hub location decision, allocation decision and air service selection decision.

(2) Demand allocation problem (median problem)

When the hubs are determined, all the demand nodes are singly allocated to "home" hubs with the objective of minimizing travel cost. The input of allocation problem includes the location of hubs and other demand nodes, demand volume between all O-D pairs, transportation cost rate for both backbone and tributary network, and number and capacity of aircraft. The corresponding decisions involve allocation of demand nodes to predetermined hubs and service selection for backbone link.

(3) Service selection problem (lower problem)

Once the hubs are determined and all the demand nodes are allocated to "home" hubs, the optimal air service can be determined for each hub link directly in Ext.1, while in Ext.2 the service selection problem is actually a flow problem with the objective of minimizing air cost and under the constraint of self-owned aircraft number and capacity. So it is exclusive for *Ext.2*. The input of the flow problem is outcome of allocation problem, air cost function, self-owned aircraft number and capacity.

The description of the three sub-problems manifests that they are "hierarchical" or "nested". The lower problem includes only air service decision, while the median problem includes extra decision on allocation besides air selection decision. The upper problem-hub location problem- is actually original problem and involves all the decisions.

- **Corresponding algorithms**

As a matter of fact, the original problem is divided according to the three decisions, i.e. hub location, allocation and service selection decisions. We propose for each decision specific algorithms. One works incorporated in another, so that the original problem can be solved near optimally in hierarchical approach with optimal and near-optimal solution of lower and median problems.

(1) Air service selection decision: integer programming

When the hub location and demand allocation decisions are made, the volume through each hub link is determined. So service selection decision in *Ext.2* can be easily made with an integer programming. This decision is passed over in the basic model for there is only one type of air service without capacity constraint, and also is passed in *Ext.1* for there is no numerical constraint on each service type.

(2) Allocation decision: LS heuristics

It has been proved that the allocation decision is *NP*-hard even when hub locations are determined[354]. The travel cost of each O-D pair of demand consists of three components: (a) the travel cost from the origin to the hub, (b) the cost between hubs (if necessary) and (c) the travel cost from the hub to the destination. We resort to LS heuristics which will be discussed in detail in Sec.4.1.5.

(3) Hub location decision: GAs

The hub location decision problem is *NP*-hard so that we resort to GAs which will be discussed in details in Sec.4.1.4.

- **Overall solution process**

---

[354] This has been proved by Kara. See Kara (1999).

We propose three specific algorithms for the three decisions, which have different roles in different sub-problems. As sub-problems are hierarchical or nested, the algorithms also work in hierarchical order. In perspective of solution process, the integer programming for air service selection is incorporated into LS for allocation decision, while the LS is again embedded in GAs for hub location decision. Therefore, GAs serves as the master algorithm for the original problem.

We apply the word "stage" to describe the solution process. A uniform hierarchical framework of solution process is proposed in Fig.4-3 for both basic and extension models in this dissertation. The upper, median and lower solution stages are highlighted with light, median and dark background respectively. We divide the whole solution process into three hierarchical and iterative stages according to the three sub-problems. Each stage solves the corresponding sub-problem described above.

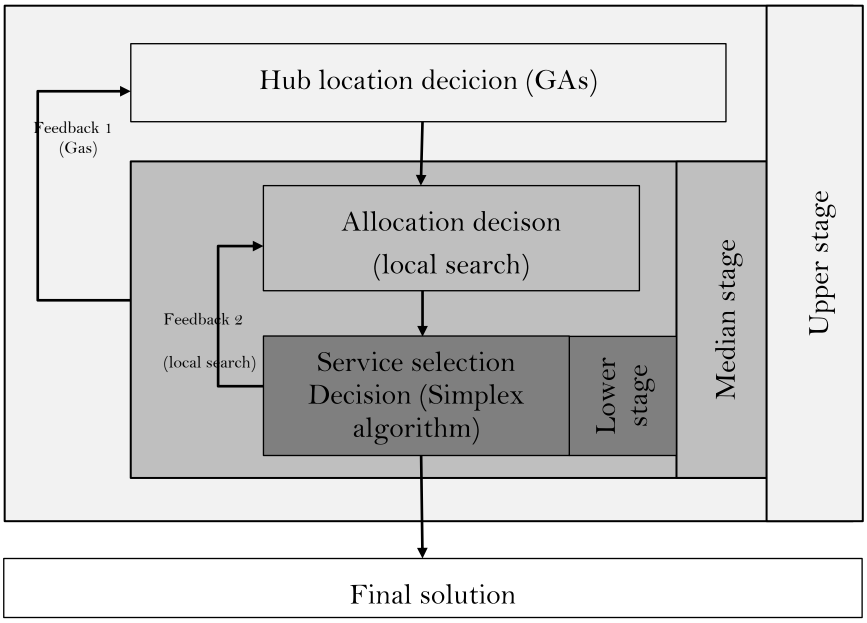

Figure 4-3: Overall solution process

In the whole solution process the three stages are interrelated with two hierarchical feedback cycles. In this sense the upper algorithms do not simultaneously solve all the underlying problems, but work cooperatively with lower algorithms. The two feedback cycles are repeated until certain stop criteria are met.

Specifically, Feedback cycle 2 is impelled by LS for allocation decision. First, the hub number and locations are fixed by GAs. On the basis of the selected hubs, demand nodes are allocated to the "home" hubs according to certain criterion. The volume through each hub link is thus determined, and the service on each hub link can be optimally determined with simplex algorithm. However, the allocation decision at allocation stage (median stage) restricts the optimal solution of service selection stage. Therefore LS of allocation decision serves as Feedback cycle 2 to seek better solution under the determined hub location.

Meanwhile, Feedback cycle 1 is impelled by GAs for hub location decision. Hub location decision at location stage restricts the optimal decision of allocation stage. We can also expect sub-optimization of the original problem if there is no iteration between allocation stage and hub location stage. GAs process takes up the role of Feedback cycle 1.

In sum, LS at allocation stage, i.e. Feedback cycle 2, attempts to minimize the travel cost based on determined hubs, while GAs at hub location stage, i.e. Feedback cycle 1, attempts to minimize the total cost of the network. With the hierarchical iteration of the subordinate algorithms the original problem is solved to near optimal finally.

4.1.4. Genetic algorithms for hub location decision

On accounts of our considerations mentioned at the end of Sec.4.1.1, we decide to take up meta-heuristics, which can yield a few near optimal solutions to the large-scale problem in a reasonably amount of time. Moreover, among dozens of meta-heuristics we adopt GAs as our master algorithms at hub location stage.

GAs is a search algorithm to find near-optimal solutions in large space. Inspired by population genetics, this idea was introduced into research field of mathematics by Holland in 1975[355]. The main thought of GAs is to employ the mechanics of natural selection to evolve a population. In the last three decades GAs has gradually become an effective and robust method for combinatorial optimization problems. There are quite a few books about GAs for reference[356].

Several factors attract us to use GAs. Firstly, GAs is widely adopted for solving various HLPs and FLPs[357]. Especially, it is a quite ideal algorithm for location problem with endogenous hub number p due to its search capability in extensive solution space, compared with other individual-based meta-heuristics. Secondly, GAs allows high degree of flexibility in the definition of the flow-dependent travel cost functions, including discrete and nonlinear functions[358], which is exactly the case in our extension model. Finally, local search for demand allocation decision can easily be incorporated into GAs, which is referred as hybrid GAs or memetic algorithm[359].

In this section, we specify customized procedures of SGAs to our problem. In Tab.4-3 we list the parameters for the algorithms in advance, also including those for improvement measures in Sec.4.2.

---

[355] See Holland (1975).

[356] See e.g. De Jong (2006); Haupt et al (2004); Gen/Cheng (2000).

[357] Besides what we have listed in Tab. 11 and 12, there are a lot more studies on GAs for p-hub median problem and p-hub center problem, e.g. Kratica/ Stanimirovic (2006), p.425; Kratica et al (2007), pp.15-28.

[358] This is also mentioned by Cunha/ Silva (2007), p.748.

[359] It has been widely applied in GAs for HLPs, such as "hub location problem with fixed cost" by Abdinnour-Helm/ Venkataramapan (1998),pp.31-51 and Cunha/ Silva (2007), pp.747-758 and hub set covering problem by Topcuoglu et al. (2005), pp.467-984.

Parameter	Description
N_{pop}	Population size
$N_{\max\_gen}$	Maximum number of generation
P_{cro}	Crossover probability
P_{mut}	Mutation probability
P_{new}	Injection rate
T_{max}	Maximum running time
N_{imp}	Maximum number of iteration without improvement

Table 4-3: Parameters for GAs

- Encoding

One solution corresponds to one individual (or chromosome in genetic terminology) in GAs. Encoding is necessary to translate the information of the solution in the form which can be recognized by GAs. And the individuals generated by GAs must be decoded into solution for us to read.

Conventional methods of encoding include binary encoding, permutation encoding, value encoding and tree encoding. Among them binary encoding is the most popular and the most feasible method for HLPs. It also requires simple procedures during the crossover and mutation operations, since the number of hubs here is variable during the reproduction.

We adopt GAs for hub location decision (not for hub location problem, see their difference in Fig.4-2 in Sec. 4.1.3). So chromosome representation includes only location decision. The length of string equals to the number of nodes in potential hub set H .The value 1 indicates that the corresponding node is selected as hub and the value 0 indicates otherwise

- Preprocessing

This is a problem-specific procedure to GAs in this dissertation due to uneven distribution of demand in China. Demand nodes locate more intensively in relatively developed East China than that in less developed West China. Or in other words, demand in West China concentrates mainly on several large cities. Some demand nodes in Western China may be so remote that they can only be covered by one potential hub or be covered by themselves if they are also in potential hub set. As we have mentioned in model assumptions in Sec.3.2.1, all demand nodes are assumed to be covered by at least one candidate hub, including itself. In view of such situation, this candidate hub is determined as hub in this preprocessing procedure and is ignored in later procedures of GAs. That is, the corresponding value in chromosome will be fixed to 1 during the crossover and mutation procedure. Meanwhile, the demand nodes that can be covered by this hub will be ignored during the initialization of solutions. But in the allocation procedure, they are treated the same as other demand nodes.

- Generating initial solutions and solution pool

In SGAs initial solutions are randomly generated. That is, initial solutions are generated with arbitrary number of hubs at arbitrary location. Solution pool is composed with a fixed number of initial solutions and is updated at each generation.

- Feasibility adjustment

Randomly generated initial solutions, as well as offspring after crossover and mutation operations, may become infeasible, i.e., some demand nodes cannot be covered by any hub. There are basically three different methods to handle infeasibility. The simplest method is to reject infeasible individuals. Nevertheless, in our problem it might be very difficult to find feasible individuals. So it is impractical to abandon all infeasible solutions. The second method is to penalize infeasible individuals a with lower fitness value[360]. The third method is to try to repair those infeasible solutions[361], which is the only method we can resort to, since the incorporated allocation problem takes feasible solution as a premise.

Feasibility adjustment in this dissertation includes two phases- adding and dropping. In adding phase, for every demand node *i* that is not covered by selected hubs. We search candidate hubs *k* that can cover it. As we can anticipate, there are at least two candidate hubs conforming to this condition, since there is preprocessing procedure beforehand. We designate the node *k* with the largest value of the following index (see Eq.4-1) as hub.

$$I_k^i = \sum_{j \in N} (w_{ij} + w_{ji}) y_{ik} / d_{ik} \qquad y_{ik} = \begin{cases} 1 & d_{ik} \le D \\ 0 & otherwise \end{cases} \tag{4-1}$$

However, this procedure is quite likely to bring redundant hubs in the solutions. So the adding phase is followed by a dropping phase with PLS, which tries to drop redundant hubs if possible. For every added hub, we drop the hub that is geographically nearest to it unless that hub is fixed during the reprocessing procedure. Then we reallocate the demand nodes formerly covered by the dropped hub to other hubs without violating feasibility. No matter whether a current hub can be dropped or not, the drop phase terminates and the algorithm goes back to adding phase for next uncovered node until all demand nodes are covered by at least one hub.

- Fitness function

A value is necessary for every feasible solution to measure its fitness to survive to the next generation. According to Darwin's evolution theory, an individual has more chance to be selected for reproduction if it is fitter.

This can be objective value of the problem, but it is not always practicable or rational. When the value of objective function is negative or unevenly distributed (either widely spread or convergent) or when there is no objective function, the fitness function is thereby adopted to evaluate the fitness of individual. It has been pointed out that the performance of GAs can be substantially improved if we use fitness function, i.e. use

---

[360] See e.g. Michalewicz/ Schoenauer (1996), p.1; Runarsson/Yao (2000), p.284.

[361] See e.g. Chootinan/Chen (2006), pp. 2263-2281; Osman et al. (2004), pp.391-405.

$f(j(x_i))$ instead of $j(x_i)$ to represent fitness of individual, where $j(x)$ can be objective value or other measurement and $f(x)$ is fitness function[362]. Conventional fitness functions include linear scaling $f(x)=ax+b$, power law scaling $f(x)=x^a$ [363] and exponential scaling $f(x)=\exp(-\beta x)$ [364].

The fitness function applied in this dissertation is based on linear scaling of the total cost derived from the objective function of the model (see Eq. 4-2).

$$fitness=\begin{cases} 0 & if\ \cos t > zerofit \\ 1-\cos t/zerofit & otherwise \end{cases} \tag{4-2}$$

The parameter $zerofit$ serves as an upper bound that eliminates some weak solutions from the process. There is no well-acknowledged approach to determine an appropriate value for the $zerofit$ parameter[365]. In this dissertation $zerofit$ is set with the maximum objective value among initial solutions.

- Selection operation

The selection criterion stipulates the method to choose parents for reproduction according to the fitness value. The criterion should comply with Darwin's evolution theory that the better the chromosomes are, the more chances they have to be selected for reproduction.

Many methods, both deterministic and probabilistic, to select the best chromosomes are available. Deterministic method is also called truncation selection, which just eliminates the weakest candidates at the end of each generation. In probabilistic selection there is still small chance for weaker solutions to survive in the selection procedure, bringing more diversification to GAs. Conventional selection methods for GAs include roulette wheel selection, tournament selection, rank selection, and so on.

We adopt roulette wheel selection in this dissertation. The idea behind this method is to allocate pie-shaped slices on a roulette wheel to each individual in the population, with each slice proportional to the corresponding individual's fitness value. Choosing individuals for reproduction in a population can be viewed as a spin of the wheel. The slice where the pointer stops is the winning one[366]. That is, if f_i denotes the fitness value of individual i in the population, its probability of being selected for reproduction is $p_i=\frac{f_i}{\sum_{j\in N_{pop}} f_j}$, where N_{pop} is population size. In this regard, the fitter individuals have greater chance to survive than the weaker ones. Here we can find the important effect of fitness function for this selection operation, for fitness function with Eq.4-2 standardizes and evens the selection probability.

- Crossover operation

[362] See Kreinovich et al. (1993), p.9.

[363] See Goldberg (1989), ch.4.

[364] See Sirag / Weisser (1987), pp.116-122.

[365] See Lim et al (2000), p.251.

[366] See Abdinnour-Helm (1998), p.491.

Crossover and mutation are the most important reproduction procedures of GAs. More specifically, crossover helps to speed up convergence rates for the search of GAs, while mutation diversifies the population. In this respect, crossover and mutation operation are two major procedures for GAs and always take place one after another (see Fig.4-4)[367]. The parameters that control how many individuals/ how much percentage of individuals undergoes crossover and mutation operation are Crossover Probability P_{cro} and Mutation Probability P_{mut} .

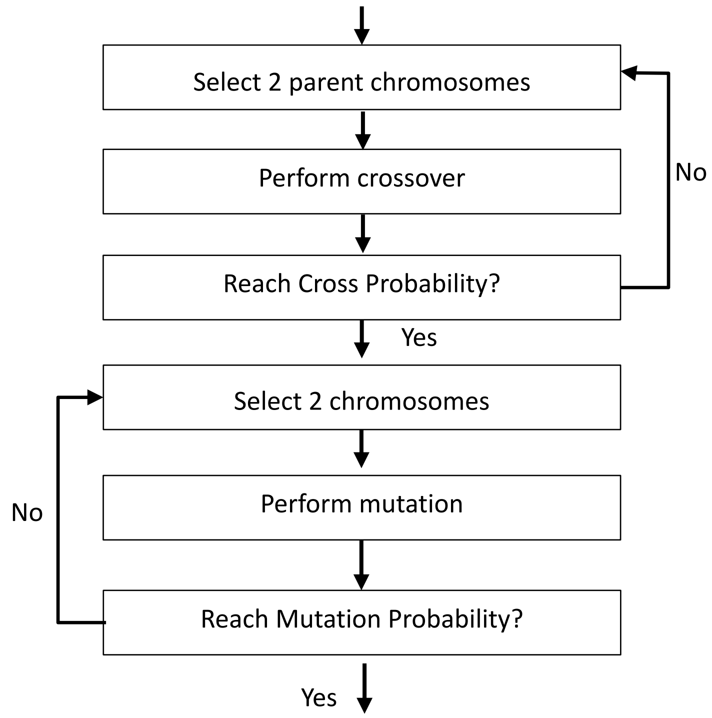

Figure 4-4: Reproduction procedure

(source: based on Lim et al (2000), p.260)

In nature environment crossover exchanges corresponding genetic material from the two parents, allowing good genes on parents to be combined in their offspring. One-point, two-point and uniform crossovers are three commonly used crossover methods. In one-point crossover one crossover point is randomly chosen. Everything after this point is exchanged in the two offspring, while everything before this point is kept the same as their parents. In uniform crossover bits are randomly copied from the first or the second parent. Uniform crossover exchanges bits of a string rather than segments of string. Studies have compared the merits of these crossover techniques both empirically and theoretically[368]. One-point and two-point crossover methods maintain low disruption rates, whereas uniform crossover is the most disruptive. Empirical evidence suggests that uniform crossover is more suitable for small population size, while the less disruptive crossover methods are better for large population size.

---

[367] In other words, crossover operation can be carried out either before or after mutation operation.

[368] See e.g. Song/ Chou (1997), p.288.

We use two-point crossover technique in this dissertation, which may be more exploratory than the one-point crossover when the population converges. Since the number of hubs is not fixed in GAs, we can randomly choose two points and exchange the segments in the middle (see Fig.4-5).

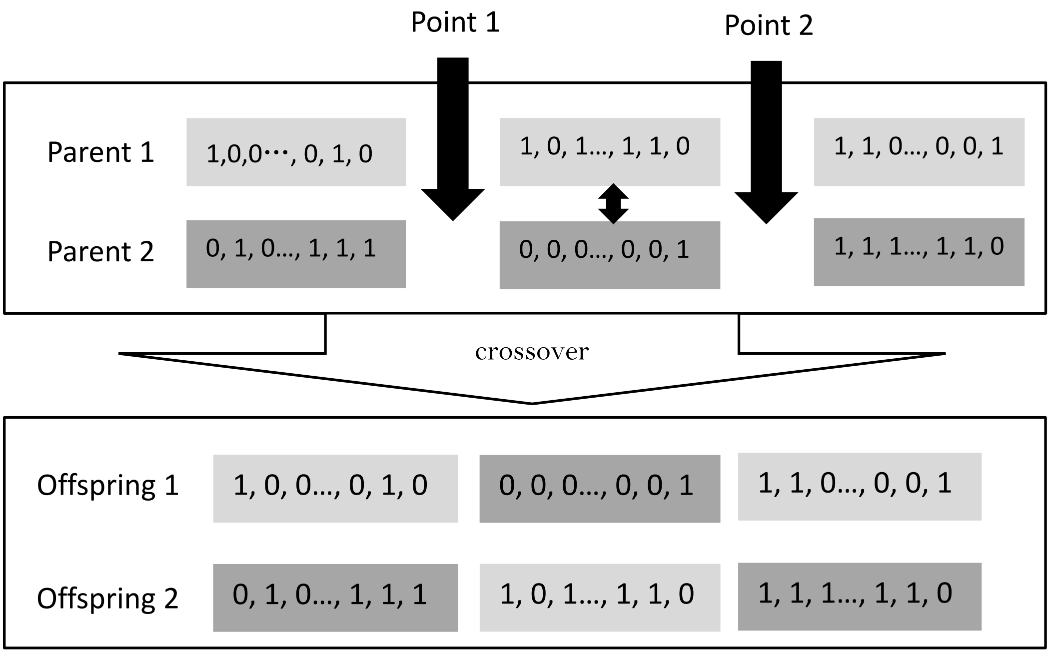

Figure 4-5: Two-point crossover procedure

- Mutation operation

After crossover, mutation procedure changes bit(s) of the chromosomes randomly. The aim of mutation is to increase diversification in the population so that search space is spread out and local optimal is prevented. First, the mutation target is selected with the same criterion as that for crossover operation. One bit of the target chromosome is then randomly chosen. If it is 0-bit, we turn it into 1-bit; and if 1-bit, we turn it into 0-bit.

- Update – Plus Strategy

At the beginning of the GAs N_{pop} individuals are randomly generated and dropped into solution pool. The solution pool will be updated later at each generation. There are two update strategies, i.e. Plus Strategy and Comma Strategy. With the former strategy, the descendant population is selected from aggregate set of incumbent generation and the newly generated offspring. With the latter strategy, the descendant population is composed merely with incumbent offspring, which requires enough offspring to put some selective pressure on the update[369]. Theoretical studies have proved that Comma Strategy is totally inefficient when the number of

[369] See Gottlieb/ Kruse (2000), pp.415-421; Jägersküpper/ Storch (2007), p.1.

offspring is not large enough, whereas the two strategies behavior equivalently with large number of offspring[370].

Since it is also quite time-consuming to repair infeasible solutions, we adopt Plus Strategy here (see Fig.4-6). Another reason to adopt Plus Strategy is that if a new generation is merely composed with newly-generated offspring, it is likely to lose the best solutions in the incumbent generation. However, best individuals (elite individuals) in the incumbent generation should be reserved for the descendant population, which is called Elitism Strategy or Steady-State Selection. In other words, some individuals can be passed over for the update procedure and go directly into the next generation. For this reason, Elitism Strategy is a method to speed up convergence toward optimum[371].

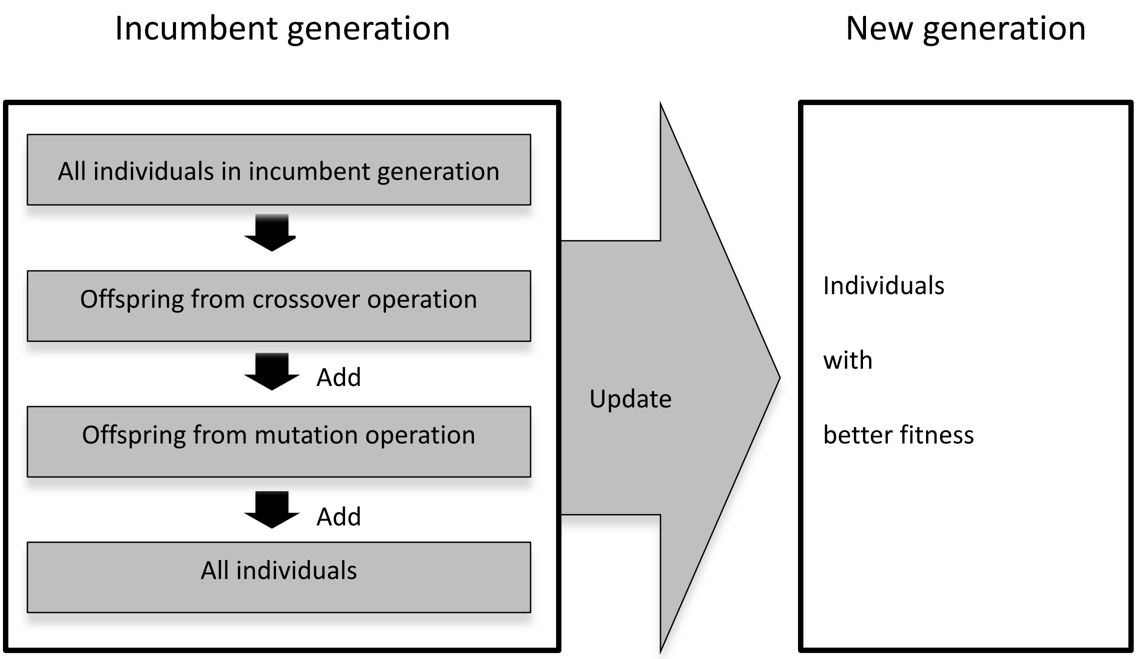

Figure 4-6: Update procedure with Plus Strategy

- Termination criterion

Improvement heuristics[372] always require a termination criterion to stop the algorithm at proper time, i.e. when search space falls into a local optimum. Only minor improvements in the solution quality can be expected when prolonging the runs. So the most widely used termination criterion of GAs is Maximum Generation $N_{\max\_gen}$. It is often associated with criterion of Maximum number of iteration without improvement, for example "no better solution or no more than 0.001% improvement after 5 iterations".

[370] See Jansen et al (2005), p. 440; Jägersküpper/ Storch (2007), p.1.

[371] See Jozefowiez et al.(2009), p.761.

[372] Heuristics process can be classified into three categories, i.e. constructive process, improvement process and incomplete exact process. See Domschke (2010), Chap.1.3; Domschke/ Drexl (1998), p.120.

The customized overall process of SGAs is illustrated in Fig.4-7. The allocation procedure will be discussed in detail in Sec.4.1.5.

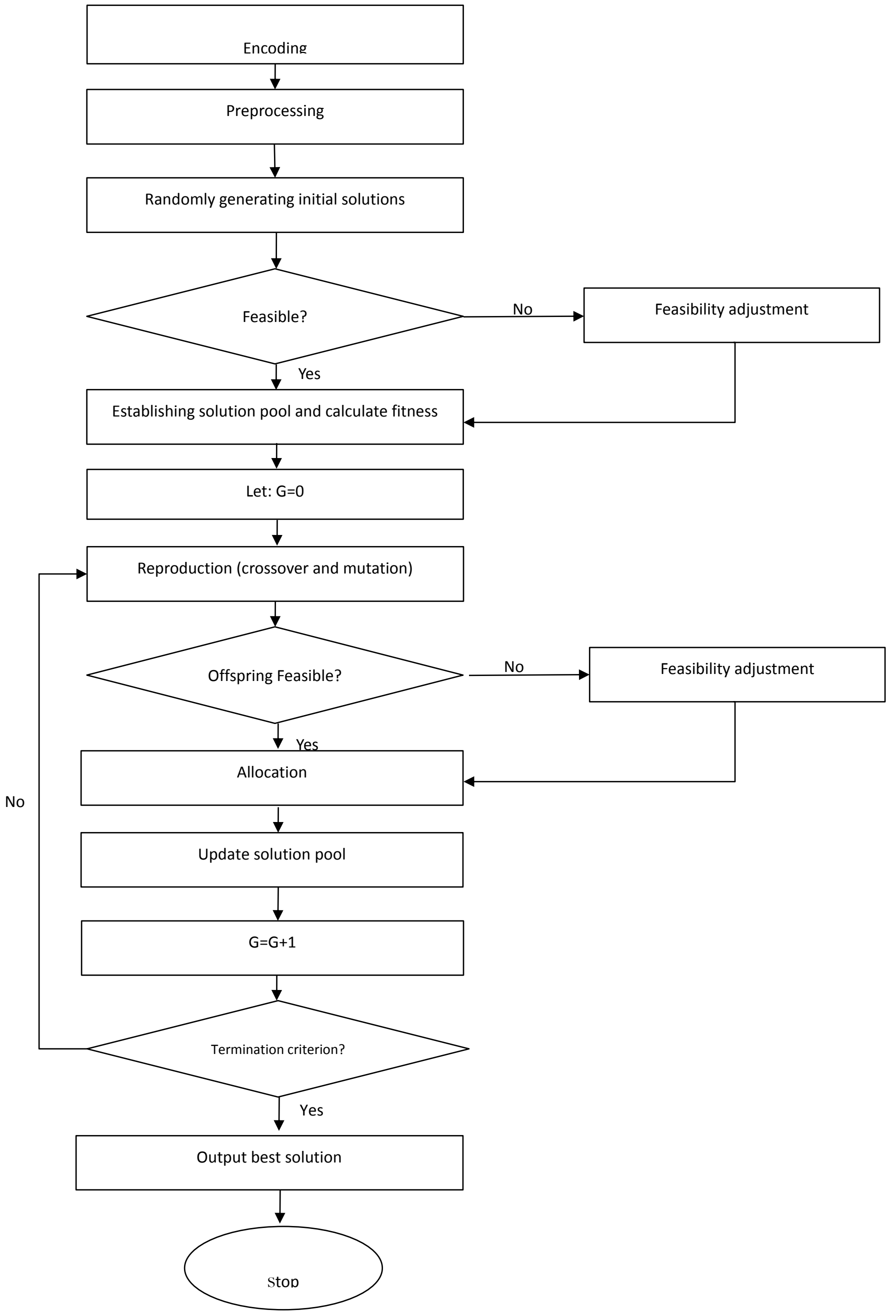

Figure 4-7: Customized overall procedure of SGAs

Like other meta-heuristics, the above-described SGAs can be easily combined with other meta-heuristics, greedy algorithm or even exact methods at both collaborative and integrative levels[373], which is called hybrid

[373] Collaborative (high) level and integrative (low) level are two hybrid methods, distinguishing the level or strength, with which different algorithms are combined. High-level combinations in principle retain the individual identities of the original algorithms. Algorithms exchange information, but are not part of each other. On the contrary, algorithms at low-level combinations strongly depend on each other. One technique is a subordi-

GAs. We incorporate subordinate algorithms into GAs for the sub-problems. Moreover, other heuristics will also be combined with GAs to improve the overall performance, which will be discussed in Sec.4.2 and be computationally tested in Chapter 5.

4.1.5. Local search for allocation decision

In multi-allocation HLP, when hubs are fixed, the allocation problem can be solved as the shortest path problem[374]. But in case of single allocation, when hub location is determined the allocation problem remains *NP*-hard[375]. Rather than simply following the least cost path, single allocation problem has to consider simultaneously three components of travel cost for each O-D pair demand, i.e. the cost from the origin to the hub, the cost between hubs (if necessary) and the cost from the hub to the destination.

The demand allocation problem is always studied as a sub-problem of HLP. In the beginning the allocation problem was treated roughly with heuristics for allocation pattern. The first two heuristics for the allocation problem were proposed by O'Kelly[376]. One of these, *Heur 1*, assigns all non-hub nodes to their nearest hubs, whereas the other, *Heur 2*, evaluates the objective value of both the nearest allocation and the second-nearest allocation before determination. In a discussion of O'Kelly's heuristics, Klincewicz[377] developed a new heuristics that uses a multi-criteria distance and flow-based allocation procedure to determine the allocation of nodes to hubs rather than relying on distance alone. This enhanced allocation rule recognizes the importance of flows in determining allocation in single allocation HLP. Given the hub locations, Campbell[378] proposed two methods for the single allocation. The first method assigns a node to the hub, through which it has its maximum flow, whereas the second allocates a node to a hub so that total transportation costs are minimized. It was shown that these two heuristics perform better than both O'Kelly's distance-based allocation pattern and Klincewicz's multi-criteria allocation pattern, and the latter method consistently provides a tighter bound then the first one.

Klincewicz[379] enhanced the algorithm for allocation by following the nearest-distance allocation pattern with exchange method. This is a great progress for the solution of allocation problem, since allocation pattern alone is destined to yield suboptimal solution. LS is followed after to correct this bias. Later more advanced techniques and algorithms have been applied to allocation LS, such as TS[380] and VNS[381]. It is believed that the quality of allocation solution depends on both starting point-the allocation pattern and the LS heuristics.

nate or embedded component of the other technique. Thus, there is a distinguished master algorithm, and at least one integrated algorithms. See Talbi (2002), p.543; Puchinger/ Raidl (2005), p.42.

[374] See e.g. Klincewicz (2002), p.112.

[375] This has been proved by Kara. See Kara (1999).

[376] See O'Kelly (1987), pp.393-404.

[377] See Klincewicz (1991), pp.25-37.

[378] See Campbell (1996), pp.923-935.

[379] See Klincewicz (1992), pp.283-302.

[380] See e.g. Calik et al (2009), pp. 3088-3096.

[381] See e.g. Ljubic (2007), pp.157-169.

However, the improvement from LS usually comes at the cost of longer computational time. Since allocation solution stage must be invoked by every individual in each generation, allocation LS procedure must be efficient in terms of time. Two measures are adopted here, i.e. Improvement Index and PLS, to improve the time efficiency.

(1) Improvement Index

LS procedures improve the solution with small moves in the predetermined neighborhood. Each small move requires the complete evaluation of the newly-generated chromosome, resulting in tremendous increase in computational time especially in large problems. It is therefore desirable to calculate the change of the objective value rather than objective value itself to accelerate the computation speed[382].

Such computational saving depends on the neighborhood structure of the LS. The most commonly used neighborhood structures for allocation include such as add-swap neighborhood, 2-swap neighborhood, and the Permutation. We adopt shift moves[383], which switch the allocation of multi-covered nodes from one hub to another. As long as the new solution achieves lower cost than the former one, the shift move is adopted. S_{hk}^{i} in Eq.4-3 refers to an Improvement Index for the cost change of the move. When demand node i is reallocated from hub h to hub k, it is calculated to determine whether the switch results in an improved solution or not.

$$S_{hk}^{i} = \sum_{j \in N} \gamma (w_{ij} + w_{ji})(d_{ki} - d_{hi}) + \sum_{l \in H} \sum_{j \in N} \beta (w_{ij} + w_{ji}) x_{jl} (d_{kl} - d_{hl}) \tag{4-3}$$

$S_{hk}^{i} < 0$ implies that the new solution has a lower cost and should therefore replace the former one, i.e. node i will be reallocated from hub h to hub k. Otherwise, the move will be ignored and LS goes on to the next move.

(2) Partial local search (PLS)

Another measure to balance between computational time and solution quality is to make distinction on how thorough the LS is to be carried out. The procedure of LS can be terminated once an improved solution is found (called first-improvement LS) or it can be repeated as many times as possible until no further improvement is possible (best-improvement LS). These are two typical examples of PLS and FLS respectively[384]. FLS puts emphasis on performing thorough search within the neighborhood of the explored solution. With each LS, the current solution is replaced by a better solution if it exists. The procedure is repeated until there is no better solution. It implies that the procedure can find the best solution in the localized area around the solution that is being explored. Compared with FLS, PLS has the advantage of simplicity and owns more potential in terms of flexibility. It is less computationally demanding so that it is especially suitable for complicated situation. With regards to this, LS procedure can be carried out more frequently in order to spread out the search by exploring many small-localized regions, thus reducing the likelihood of the algorithm being trapped in a local optimum.

---

[382] See e.g. Hansen/ Mladenovic (1997), pp.207-226; Resende / Werneck (2007), pp.205-230.

[383] See Skorin-Kapov et al. (1996), pp. 582-593.

[384] See e.g. Blum/ Roli (2008), p.1-30; Lim/ Omatu (2000), p.258.

PLS is adopted here for allocation decision mainly out of two reasons, namely computational time and improvement impact. For one thing, LS procedure for allocation is invoked by every individual at each generation. When population size is 50, this procedure will be called at least 2500 times with termination criterion of at least 50 generation. It must be time efficient. For another, optimization of allocation decision can make little advance to the solution of the overall problem unless the hub location decision is optimal in advance.

4.1.6. Integer programming for service selection decision

In order to solve *Ext.2*, the solution process follows the process framework in Fig. 4-3 in section 4.1.3. We have introduced the algorithm for hub location decision and allocation decision in section 4.1.4 and 4.1.5 respectively. In this section, we introduce the algorithm for service selection decision.

After hubs are selected and all the demand nodes are allocated to "home" hubs, the inter-hub flow is determined. We denote it as w_{km}, where $w_{km} = \sum_{i \in N} \sum_{j \in N} x_{ijkm} w_{ij}$, $x_{kk} = 1$ and $x_{mm} = 1$. Air service is selected for each hub link with the objective of minimizing air cost and under the constraint of aircraft number and capacity. Since the backbone network is fully connected by direct flight, the remaining problem becomes an integer programming. The Integer Programming Toolbox in Matlab is applied and embedded in our algorithm.

$$\text{Minimum} \sum_{k \in H} \sum_{m \in H} \sum_{p \in P} SF'_{km}(w_{km}) \tag{4-4}$$

$$\text{S.T.}\ u_p^l z_{km}^p < w_{km} \quad \forall k, m \in H,\ k \neq m,\ p \in [1, 2 \ldots, P] \tag{4-5}$$

$$w_{km} \leq u_p^u + (1 - z_{km}^p) M \quad \forall k, m \in H,\ k \neq m,\ p \in [1, 2 \ldots, P] \tag{4-6}$$

Also constraints (3- 25), (3-28) and (3-29)

4.2. Improvement techniques

The relationship between all the improvement techniques proposed in this section and customized SGAs introduced in Section 4.1.4 are illustrated in Fig.4-8. Improvement techniques 1 and 4 are integrative combination with meta-heuristics, while Improvement techniques 3 belongs to cooperative combination with meta-heuristics (see Fig.4-1).

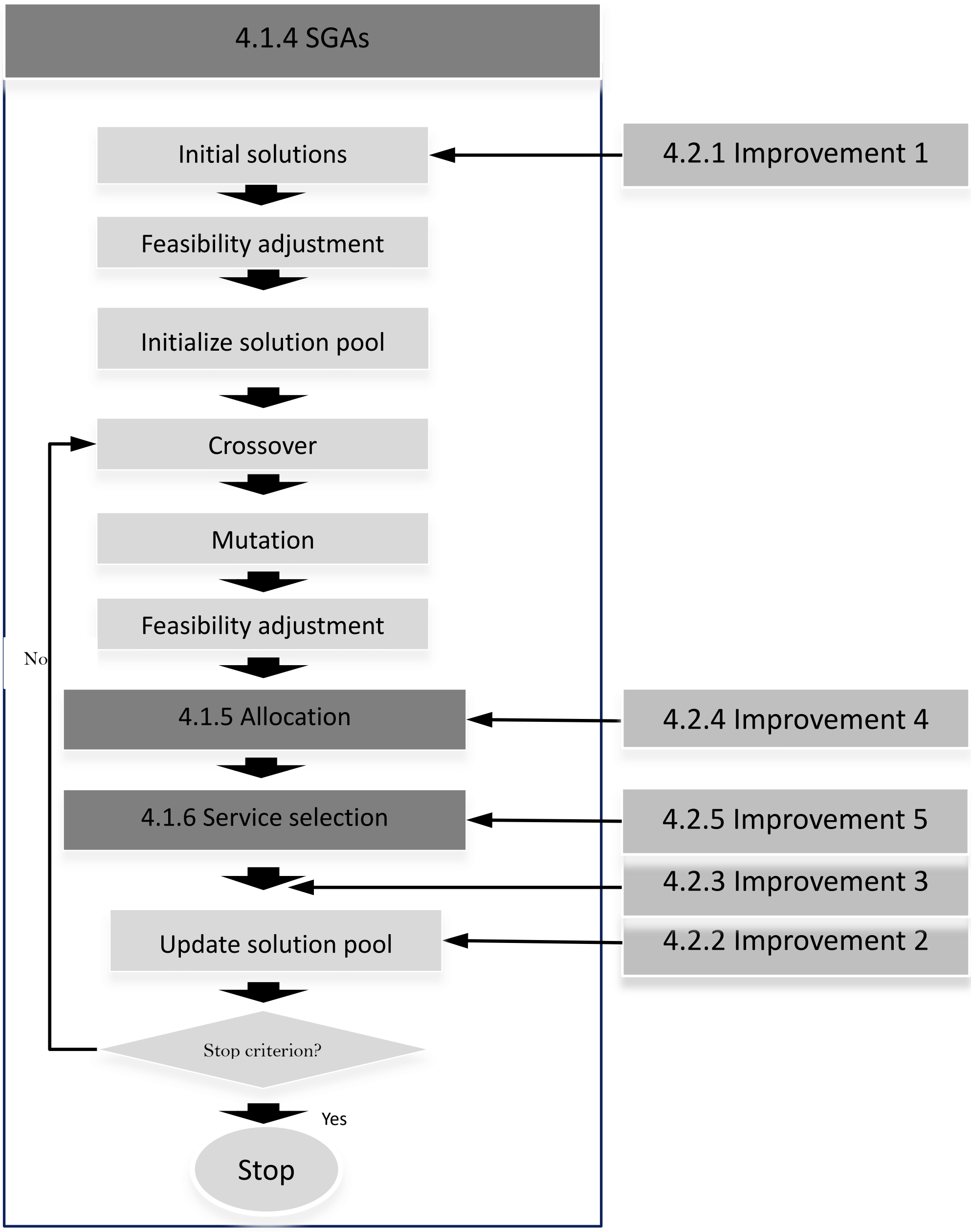

Figure 4-8: Relationship between SGAs and improvement techniques

4.2.1. Improvement 1: constructive procedure for initial solution generation

The choice of starting point is very crucial for the performance of iterative improvement heuristics such as GAs. In this section, we try to improve the performance of GAs with a "good" initial population.

It is common and easy for SGAs to randomly generate initial population for HLPs, if the hub number p is exogenously fixed. However, for HLPs with endogenous hub number, such as hub set covering problem or hub location problem with fixed cost, it is quite unreasonable and arbitrary to choose random number of hubs at random place from the candidate hub set H. For the hub set covering problem, the primary problem is the feasibility of the solutions. Randomly generated solution has low possibility that all demand nodes can be covered by the selected hubs, so that the program spends a lot of time to judge the feasibility of the new generated solution and to repair the unfeasible solutions. For hub location problem with fixed cost, the primary problem is too large search space, so that the program is likely to run randomly and yields pool solution even after many generations. Moreover, randomly generated solutions are also not applicable for network with unevenly spread demand nodes, which is exactly the situation in our case.

In the light of these problems, we replace the random procedure for generating initial solutions in SGAs with "constructive procedure" borrowed from GRASP, by which feasible solutions can be generated. We name GAs incorporated with "constructive procedure" Constructive GAs (CGAs) in the following. This "constructive procedure" is iterative, greedy, random and adaptive. It is iterative because every initial solution is constructed by choosing one element at a time until the solution is feasible. It is greedy because the addition of each element is guided by the myopic criterion or greedy function. It is random because every element is chosen randomly from a Restricted Candidate List (RCL). The RCL technique allows the procedure to be repeated but to generate different initial solutions every time. Finally, it is adaptive because the RCL is updated after each choice.

An appropriate composition of RCL signifies a good balance between the diversity and the intensity of initial population. The RCL is constructed by repeatedly calculating the greedy function $f(k)$ (see Eq.4-7) and ranking the results in decreasing order. Actually the function $f(k)$ denotes the distance-weighted flow of those still uncovered nodes that hub *k* can cover. This greedy function is based on the assumption that the busier a hub is and the nearer a hub to demand nodes is, the more value it will bring to the network and the more travel cost it will save.

$$f(k) = \sum_{i \in N_{crasp}} \sum_{j \in N} (w_{ij} + w_{ji}) y_{ik} / d_{ik} \qquad k \in H_{crasp} \qquad y_{ik} = \begin{cases} 1 & d_{ik} \le D \\ 0 & otherwise \end{cases} \tag{4-7}$$

In the beginning, the to-be-assigned hub set H_{crasp} contains all potential hubs in H; the to-be-allocated demand node set N_{crasp} contains all elements in N. The to-be-constructed location solution is ϕ. We calculate the objective value of $f(k)$ for every $k(k \in H_{crasp})$. Then we assign $f^{\max}$ and $f^{\min}$ with the largest and the smallest values of $f(k)$ respectively. We also define a threshold parameter $\alpha(\alpha \in [0,1])$. The RCL is composed of all elements in H_{crasp}, whose objective values of $f(k)$ are superior to the threshold, i.e. $\left[f^{\min} + \alpha(f^{\max} - f^{\min}), f^{\max} \right]$. After one node is randomly and uniformly selected from the RCL as hub node, it is deleted from H_{crasp} and added in the location solution. At the same time all demand nodes in N_{crasp} that can be covered by this selected hub are eliminated from N_{crasp}. So H_{crasp}, N_{crasp}, $f^{\max}$, $f^{\min}$ and the location solution are updated after each selection. This procedure goes on until all demand nodes are covered, i.e. N_{crasp} is ϕ (see Fig.4-9). It will be repeated $|N_{pop}|$ times to generate enough initial solutions for the solution pool.

As we can see, this constructive procedure can not only yield feasible solutions, but also control the randomness of the initial population. Borrowed from GRASP, it can be regarded as a repetitive sampling technique. Each reiteration produces a sample from a distribution, whose mean and variance are dependent on the parameter α . The parameter α controls the greediness and randomness or, in other words, intensity and diversity of the population. The case $\alpha = 1$ corresponds to a pure greedy algorithm, while the case $\alpha = 0$ corresponds to a pure random algorithm.

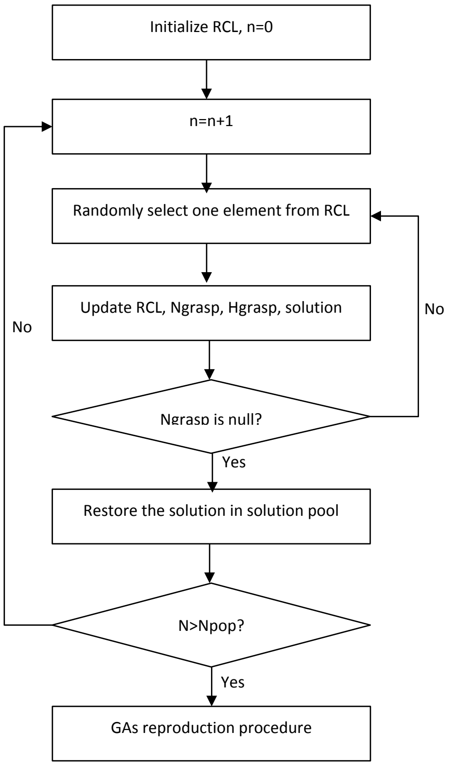

Figure 4-9: Constructive procedure for initial solution generation

4.2.2. Improvement 2: injection mechanism

The injection mechanism is to introduce new chromosomes into the population at every generation or at regular generation intervals. Contrary to the constructive procedure that controls the diversity of the initial population, the injection mechanism controls the diversity of the population during the GAs process. Some new generated individuals are injected into the solution pool to increase the diversity of the population throughout generations and to prevent premature of the algorithms. This idea is borrowed from the natural phenomenon, simulating the immigration of people between countries or regions. This scheme has been proved effective when embedded in GAs to solve location problems[385].

The parameter P_{new} is applied to control the diversity of the population during the GAs process. In order to keep the population size constant, $(1-P_{new}) \times |N_{pop}|$ best individuals are selected from the former generation with "Plus update strategy", together with the new generated individuals, to compose the next generation (see Fig.4-10).

[385] See Salhi/ Gamal (2003), p214.

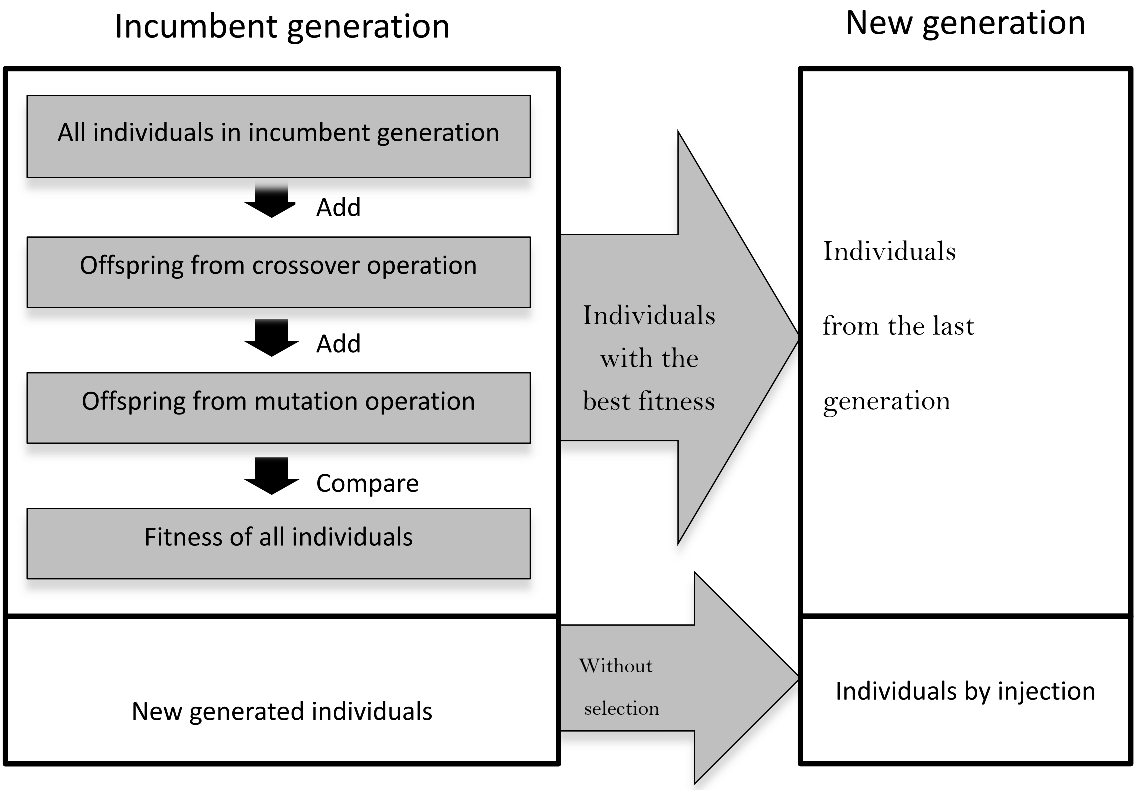

Figure 4-10: Plus update strategy with the injection mechanism

4.2.3. Improvement 3: local search after GAs

As a population-based algorithm, SGAs is good at quickly locating the regions with high performance in vast and complex search space. Once those regions are located, it may not be able to explore the complex space sufficiently[386]. That is to say, it is relatively weak in exploiting the regions that have been found. It is believed that GAs can be more efficient (i.e. need less time) and more effective (i.e. find better solutions) for most combinatorial optimization problems when embedded with LS techniques[387]. One possible reason given by Jaszkiewicz[388] is that in many cases local optimum constitutes a relatively small part of the search space and thus can be achieved in an efficient way. LS provides the potential to cover the weakness of GAs in searching local areas after GAs efficiently find the vicinity of the optimal solution from a wide range.

GAs, which use certain kind of interaction with local searchers, are named Memetic Algorithms (MAs)[389]. The name MAs is inspired by Richard Dawkin's concept of a meme[390], which in GAs refers to the strategies (e.g.

[386] This may be one of the reasons why the performance of SGAs deteriorates significantly as the size of the problem increases. See Lim et al (2000), p.249.

[387] See e.g. Merz /Freisleben (1997), pp.159-164; Galinier /Hao (1999), pp.379-397; Jaszkiewicz (2002), pp.50-71; Goldberg/ Voessner (1999), pp.220-228; Galinier/ Hao (1999), pp.379–397.

[388] See Jaszkiewicz (2002), pp. 50–71.

[389] It was so named firstly by Moscato (1989), p.2003;

[390] See Dawkins (2006), p.189.

local refinement, perturbation, etc.) that are employed to improve individuals. It combines the evolutionary adaptation with individual learning during the whole process. That is, the memes reshape the search space and they themselves can adapt to the reshaped space. In different contexts and situations, MAs are also called Hybrid GAs[391], Genetic Local Searchers (GLS)[392], Baldwinian GAs[393] or Lamarkian GAs[394].

LS starts from a given initial solution as the current solution and checks its neighborhood for a better solution. If such solutions exist, the algorithm replaces the current solution with the best solution found in the neighborhood and repeats this procedure. In case the LS cannot find solution better than the current solution in the neighborhood, the algorithm returns the current solution and terminates. This method does not guarantee globally optimal solutions if it does not search as fully as enumeration. But most of time it returns relatively good solutions.

The effectiveness of LS depends on several aspects, such as the neighborhood structure, search technique, the evaluation of neighbors, and its starting point. To this end we are faced up with the problem how to tradeoff between computational time and effectiveness? This problem results in a series of questions: Where and when should LS be invoked? Which individuals in the population should be improved by LS? How much computational effort should be allocated?

In this dissertation LS is attached after GAs to improve the final solutions of GAs, i.e. we apply LS on both hub location decisions and allocation decisions to each individual in the final solution pool of GAs. That is to say, the LS works collaboratively with GAs[395].

- LS for hub location decisions

The LS is applied for hub location decisions to check if the final solutions from GAs can be further improved by LS. We use the 2-swap neighborhood for the LS, since it always yields the best result in short time[396]. It is also named single-relocation algorithm[397]. The algorithm looks for a pair of hubs: one to be inserted into current solution, the other to be removed. If the new solution is not feasible, we just delete it and find another pair. When a feasible hub location solution is found, the remaining demand allocation problem is solved with the same process as illustrated in Fig.4-3 in Sec.4.1.3. The fitness of this solution is then calculated and compared with the incumbent best solution to check if it needs to be updated. Finally, the LS returns the best solution.

Even after the neighborhood structure of LS is determined, various tactics can be employed in this context to tradeoff between the search depth in the neighborhood space and the computational resource. Accounting for the good performance of GAs in former researches, we use FLS, which tries to improve the target solution

[391] See He/ Mort (2000), p.42; Vazquez/ Whitley (2000), p.135; Fleurent/ Ferland (1994), p.173; Lim/ Omatu (2000), p.258.

[392] See Merz (2000).

[393] See Ku/ Mak (1998), pp. 481-490.

[394] See Morris et al (1998), p.1641.

[395] See Sec. 4.1.2.

[396] See Resende/ Werneck (2007), p.207.

[397] See Pamuk/ Sepil (2001), p.402.

until all hub pairs have been checked. However, FLS does not mean enumeration, since we only explore one specific neighborhood.

- LS for allocation decisions

We adopt PLS for allocation decisions during GAs process under the consideration of computational time and effectiveness to the final solution[398]. Here we adopt FLS for allocation decision after GAs for the same considerations. For one thing, this process is invoked only one time after GAs rather than at the end of every generation of GAs. For another, there is still potential to further improve the allocation decision even after PLS for each generation. Therefore, a FLS for allocation decision is followed after GAs, i.e. after LS for hub location decision, so that the updated solutions from FLS for hub location decision can be further improved by allocation FLS (see Fig.4-11).

We still use shift moves[399], which switch the allocation of multi-covered nodes from one hub to another until all the possibilities have been considered and no improvement could be made. Improvement Index of S_{hk}^{i} (see Eq.4-3) can still be applied to save computational resource.

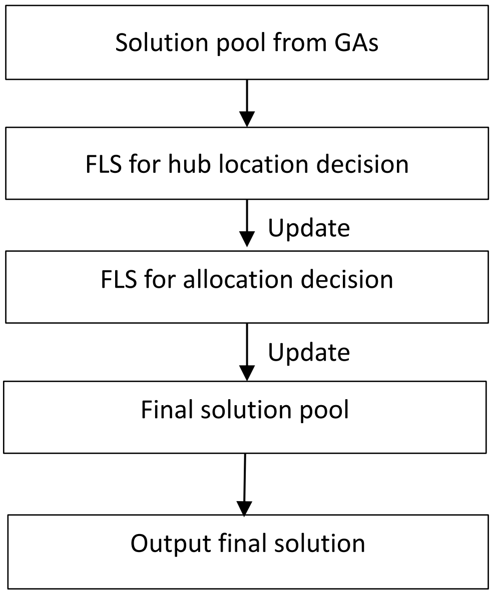

Figure 4-11: Full local search after GAs

4.2.4. Improvement 4: initial solution for allocation local search

This improvement technique is aimed at providing a relatively "good" starting point for LS for allocation decisions at each generation, which is studied in Sec.4.1.5. Since the strength of LS is to exploit better solution in a relatively small solution space, a "good" starting point is essential to the result of LS[400].

[398] For details, please see Sec.4.1.5.

[399] See Sec. 4.1.5.

[400] See Blum/ Roli (2008), p.7.

According to the literature review on allocation decision in Sec.4.1.5, the most commonly used allocation patterns (or allocation criteria) include “distance-based” allocation, “multi-criteria” allocation, “maximum flow” allocation and “minimum cost” allocation. Except the first one, all the allocation patterns are constructive heuristics, by which the elements in the solution are added one by one[401]. Actually, the right choice of allocation pattern for different HLPs needs specific knowledge of the problems.

(1) Distance-based allocation pattern by O’Kelly[402]

This allocation pattern, including both “nearest distance” and “second-nearest distance” allocation, was initially proposed by O’Kelly, who pointed out later that when the “nearest-distance” allocation criterion is always the best for incapacitated *p*-median problem, it is not always true for interacting facility location problems, since it ignores the flows among facilities[403]. However, later studies on HLPs find that the “nearest distance” allocation pattern is still an effective method, especially for p-hub median problem, under the consideration of short computational time[404].

(2) Multi –criteria allocation pattern by Klincewicz [405]

A “multi-criteria” allocation pattern to the *p*-hub median problem was proposed by Klincewicz: initial allocation of a given node is based on the sum of common traffic with certain hub and the distance to that hub. We carry out this idea by allocating non-hub nodes to “home” hub according to the two steps as follows.

(a) For the non-hub node that is covered by only one hub, it will be allocated to that hub directly;

(b) For the non-hub node i that can be covered by more than one hub, Index $A1$ (Eq.4-8) will be used as the criterion for the allocation decision.

$$A1_h^i = \frac{\sum_{j \in N}(w_{ij} + w_{ji})x_{jh}}{d_{ih}} \tag{4-8}$$

This index considers not only the distance d_{ih} between node i and hub h but also the interchange volume $\sum_{j \in N}(w_{ij} + w_{ji})x_{jh}$ between node i and all nodes that have already been allocated to hub *h*. This index implies that if a node has heavier flow with a candidate hub and/or shorter distance from that hub, it is probably more economical to allocate this node to that hub.

As we can anticipate, part of non-hub nodes is covered by only one hub. After step one, each hub already has some subordinate demand nodes. For every not-yet-allocated demand node i, we calculate the Index $A1$ for

---

[401] See Mayer (2001), p.91.

[402] See O’Kelly (1987), pp.393-404.

[403] See O’Kelly (1992), p.303.

[404] See e.g. Kratica (2007), pp.15-28.

[405] See Klincewicz (1991), pp.25-37.

all hubs h that can cover it. We allocate the demand node to the hub with the highest index value. Every double- or multi-covered non-hub node is under this consideration till all find their "home" hub.

(3) Maximum flow allocation pattern by Campbell

The success of multi-criteria allocation pattern implies that traffic volume to potential "home" hub could also serve as a criterion for allocation. Travel cost by air between hubs is normally higher than that by truck, making this idea more attractive for air-ground network than multi-criteria allocation pattern. The corresponding index for the to-be allocated non-hub node i is as Eq.4-9.

$$A2_h^i = \sum_{j \in N} (w_{ij} + w_{ji}) x_{jh} \tag{4-9}$$

(4) Minimum cost allocation pattern by Campbell [406]

Campbell proposed another allocation pattern called "minimum cost allocation", which allocates a node to a hub so that total travel costs are minimized. It was shown that this method consistently provides a tighter bound then the above one. The corresponding index for the to-be allocated non-hub node i is as Eq.4-10.

$$A3_h^i = \sum_{j \in N} \sum_{m \in H} (w_{ij} + w_{ji}) x_{mj} (\gamma_{ih} d_{ih} + \beta_{km} d_{km} + \gamma_{mj} d_{mj}) \tag{4-10}$$

In Chap.5 we will test all these four allocation patterns to evaluate their performance under our instance.

4.2.5. Improvement 5: approximate of integer programming in early stage

In the basic model and *Ext.1*, the demand volume on each hub link can be calculated when the hub location and allocation is determined, so that the best service type can be easily chosen directly according to the cost function of each service type. In other words, integer programming is not necessary. However, the air service selection decision in *Ext.2* is determined with an integer programming problem due to the numerical constraints on the aircraft in current fleet.

If the integer programming for the service selection decision is time-costly, the total running time will grow exponentially. This is always the bottleneck to adopt hierarchical algorithm to large-scale instances. So we propose the following method, trying to make the overall algorithm more time-efficient and keep its performance at the same time. Our method is like this: if the improvement on the solution by solving the embedded sub-problem is relatively small compared with the improvement by solving the overall problem, we use an approximate result of the sub-problem in the master algorithm rather than solving the sub-problem until the master algorithm find the near-optimal region.

Generally speaking, if the algorithm for the sub-problem is time-costly and must be invoked frequently, this time-saving method can be efficient with quite small negative impact on the performance of the overall algorithm.

[406] See Campbell (1996), pp.923-935.

4.3. Summary

This chapter is contributed to the design of solution process, individual algorithms for decisions and improvement techniques.

After we make a relatively thorough literature review on solution methods for related HLPs in Sec.4.1.1, we decide to adopt meta-heuristics under the consideration of instances scale and management requirement. In Sec.4.1.2 we briefly discuss the classification of hybrid meta-heuristics and hybrid principle. Sec.4.1.3 designs the solution process. The literature review on solution process of compound location problems demonstrates that our problem is essentially a location problem with embedded allocation and service selection problems. We divide the original problem into three hierarchical sub-problems and propose an overall solution process, connecting all subordinate algorithms with two hierarchical feedback cycles. Form Sec.4.1.4 to Sec.4.1.6 we propose specific algorithms for individual decisions. Specifically, in Sec.4.1.4 we propose the customized procedure of GAs for hub location decisions. In Sec. 4.1.5 we illustrate LS algorithms for allocation decisions. In order to balance the solution quality and computational time, we take up two measures, i.e. Improvement Index and partial LS. In Sec. 4.1.6 we use integer programming for the service selection problem with predetermined hub location and allocation decisions.

In Sec.4.2 we propose five improvement techniques for different procedures of SGAs. Improvement 1 is oriented towards initial solution generation procedure. We incorporate constructive procedure borrowed from GRASP to generate initial solutions for GAs. This method can not only yield feasible solutions but also balance the diversity and intensity of the initial solution pool. Improvement 2 is oriented towards the update strategy of the solution pool. In contrast to constructive procedure for initial solution generation, injection mechanism balances the diversity and intensity of the solution pool during the whole process of GAs. Improvement 3 tries to further improve the solutions from GAs by attaching LS after GAs for both hub location decisions and allocation decisions. Improvement 4 tries to provide a “good” initial solution for allocation LS. We list four different initialization methods. Improvement 5 is to raise the computational efficiency with no or little negative impact on the solution quality. These improvement techniques can be classified into the following categories of the hybrid heuristics according to Fig.1-4

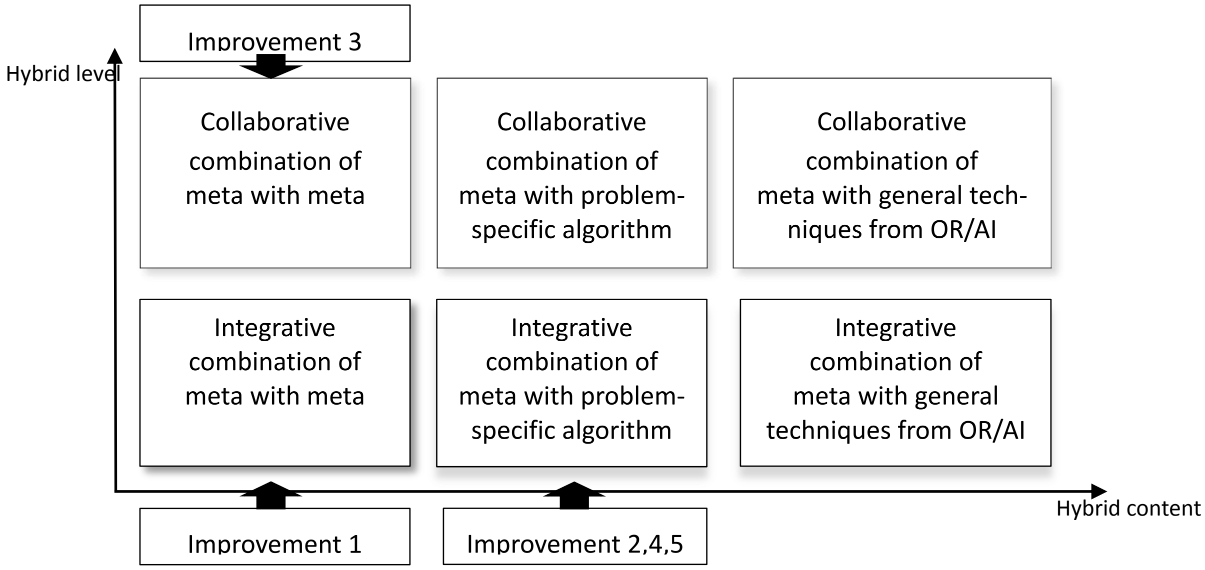

Figure 4-12: Classification of improvement techniques

5. Computational tests for algorithms

In this chapter, we test the performance of the tailored algorithms we have proposed in last chapter. Performances are evaluated in terms of solution quality, i.e. deviation from the best-known solution, and running time. All the computational tests for algorithms are based on basic models. The test results are also applicable to the two extension models under the same instance, since air service selection problem is solved to optimum if invoked. For this reason, we believe that the results based on basic model can represent the performance (in terms of solution quality) of the algorithms.

Sec.5.1 is dedicated to tests on the overall performance of the proposed hybrid GAs under small-scale instances with the CAB data set. The performance is evaluated by comparing its solutions with optimal solutions generated by CPLEX. Computational tests are carried out with various instance dimensions of hub coverage radius, hub fixed cost and instance scale.

Sec.5.2 focuses on evaluating the performance of the proposed improvement techniques under large-scale instances with the AP data set. Since the optimal solutions or even benchmark solutions for large-scale instances are not available, we test if those techniques can further enhance the performance of algorithms. We first modify test data and make preliminary tests to set parameters for GAs. Then we test whether and how each improvement technique can improve the performance of GAs. Finally, we compare the overall performance of proposed hybrid GAs with SGAs.

In all computational tests in this chapter and next chapter, we apply the following assumptions for the sake of simplicity.

(1) The twilight and sunrise regional sorting time is the same for all potential hubs. Together with the assumption for the model formulation that time window between the earliest departure and the latest arrival of aircraft is the same for all potential hubs, it denotes that the twilight cutoff time and sunrise setup time for every potential hub is the same.
(2) The coverage radius for all potential hubs is the same.
(3) The cost rate for each backbone link is the same. So is the cost rate for each feeder link.
(4) Both the fixed cost and variable cost for one type of aircraft are assumed to be the same (for Chap.6).

5.1. Tests for small-scale instances

As we have mentioned in Sec.2.4.2, up till now we have only found two similar works that study a similar model as ours. Thaddeus Kim Teck Sim, who studied a similar model as our basic model in his doctor dissertation[407], has made some progress on this problem recently[408] and provides us a series of benchmark solutions under small-scale instances generated by CPLEX.

- Description of test instances and algorithm parameters

[407] See Sim (2007), available on internet: http://ir.uiowa.edu/etd/124

[408] See Lowe/ Sim (2012).

The benchmark solutions are based on well-known Civil Aeronautics Board (CAB) data set[409]. O'Kelly first introduced the data set that is based on the airline passenger interactions between 25 US cities in 1970 evaluated by the Civil Aeronautics Board (CAB)[410]. This data set has been widely used for testing HLPs. It includes demand flow and travel distance between 25 large cities in the United States.

Lowe and Sim[411] made some specifications on the input of the computational tests. We illustrate them in the following with the parameters adopted in this dissertation. The travel cost rates in backbone and tributary network β and γ are taken as 0.75/25000 and 1/25000, respectively. Two different hub fixed cost data sets are created: "constF" and "diffF" fixed cost data sets. We only adopt the "constF" data set, which assumes that the cost of opening a hub is the same regardless of where the hub is located. Computational tests are carried out with hub fixed cost fh_k as 5000 and 50000, respectively. The coverage radius of hub in the tests is proportional to the length of the longest link in the corresponding network. Specifically, Δ in Tab.14 represents the longest links in the networks with 10, 15, 20 and nodes respectively. Hub coverage radii in our tests are taken 10% and 50% of the longest links in the corresponding networks, which are denoted as *R1* and *R2* in Tab.5-1.

	Δ	R1=0.1 Δ	R2=0.5 Δ
n=10	1764.791	176.4791	882.3955
n=15	2600.078	260.0078	1300.039
n=20	2600.078	260.078	1300.039
n=25	2725.79	272.579	1362.895

Table 5-1: Hub coverage radii under small-scale test instances with the CAB data set

Several probe runs are carried out to roughly determine the algorithm parameters for tests under small-scale instances. Preference is then given to Population size $N_{pop} = 50$, Mutation probability $P_{cro} = 0.9$ and Crossover probability $P_{mut} = 0.3$. Concerning the improvement techniques, we adopt CGA (75%)[412] with "maximum flow allocation pattern" followed by PLS during the process of GAs but without "injection mechanism". The algorithm is forced to run at least 30 generations and is terminated when there is no more than 0.001% improvement of the solution pool after 5 generations. We also make FLS of allocation decision on best solution after GAs.

The comparison with optimal solutions generated by CPLEX includes four instance scales with 10, 15, 20 and 25 nodes respectively. Under each scale, there are 4 different combinations of hub coverage radius (2 variations) and hub fixed cost (2 variations). Considering these three instance dimensions, we have totally 16 test instances.

- Stability of the algorithm

We run the algorithms 10 times under each instance. Besides the best, worst and average solutions, we also show the stability of the algorithm under these small-scale instances. In probability theory and statistics, the

[409] It is available on internet: http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/phub4.txt. （access on 20.01.2013）

[410] See O'Kelly (1987), pp.393- 404.

[411] See Lowe/ Sim (2012).

[412] 75% refers to the parameter α for constructive procedure illustrated in Sec.4.2.1.

coefficient of variation (CV) is a normalized measure of dispersion of a probability distribution. It can be applied to indicate the stability of individuals in the group. Normally, CV is defined as the ratio of the standard deviation (σ) to the mean (μ) (see Eq.5-1). The smaller the value of CV is, the more stable the individuals in the group are.

$$cv = \frac{\sigma}{\mu} \tag{5-1}$$

- Test results and analysis

The proposed hybrid GAs is coded in Matlab6.5. The execution is conducted on Intel(R) Core(TM) Duo CPU P8600 2.4GHz with 2G RAM. The benchmark solutions of HCFPs were obtained by using CPLEX 12.3 on an Intel Core 2 2.13GHz PC with 2GB of RAM.

Problem size	Radius (delta)	Fixed cost	Total cost							Calculation time (s)	
			Optimal	GAs(best)	Dev. best (%)	Dev.worst (%)	Dev.avg. (%)	CV	Optimal times	CPLEX	GAs (avg.)
n=10	0.1	5000	6.4106E+04	6.4106E+04	0.00%	0.64%	0.45%	0.0031	3	0.014411	0.9782
		50000	4.6911E+05	4.6868E+05	0.09%	0.09%	0.09%	0.0000	0	0.014082	0.9641
	0.5	5000	4.1024E+04	4.1024E+04	0.00%	13.01%	3.25%	0.0502	2	0.086582	0.1888
		50000	1.3102E+05	1.3102E+05	0.00%	4.08%	0.69%	0.0119	2	0.0812	0.1871
n=15	0.1	5000	1.2894E+05	1.2355E+05	4.18%	4.18%	4.18%	0.0000	0	0.062865	1.4476
		50000	6.2795E+05	6.2795E+05	0.00%	0.22%	0.11%	0.0012	5	0.06027	1.4661
	0.5	5000	1.1345E+05	1.1291E+05	0.47%	10.85%	2.55%	0.0426	0	0.88197	0.2449
		50000	2.1321E+05	2.1321E+05	0.00%	2.60%	0.48%	0.0076	3	0.65473	0.2464
n=20	0.1	5000	2.2945E+05	2.2431E+05	2.24%	2.24%	2.24%	0.0000	0	0.25745	1.8207
		50000	7.7755E+05	7.7047E+05	0.91%	0.91%	0.91%	0.0000	0	0.21868	1.8346
	0.5	5000	2.2008E+05	2.2008E+05	0.00%	30.07%	6.04%	0.0845	4	1.9693	0.2949
		50000	3.7041E+05	3.7041E+05	0.00%	4.42%	2.20%	0.0139	2	1.7556	0.2901
n=25	0.1	5000	3.3379E+05	3.2114E+05	3.79%	3.79%	3.79%	0.0000	0	0.70742	2.2213
		50000	9.3570E+05	9.3570E+05	0.00%	0.83%	0.33%	0.0043	6	0.73232	2.2279
	0.5	5000	3.3056E+05	3.1449E+05	4.86%	5.14%	4.97%	0.0014	0	4.7378	0.2932
		50000	5.3413E+05	5.3413E+05	0.00%	0.64%	0.32%	0.0034	6	24.357	0.2932

Table 5-2: Results of computational tests under small-scale instances

The test results, together with benchmark solutions, are displayed in Tab.5-2. In the first three columns instance's dimensions (problem size, radius parameter and hub fixed cost) are given. The overall performance of the proposed algorithms is evaluated in terms of solution quality and running time. The columns 4 to 9 contain information about solution quality, while the last two columns concern running time. The benchmark solutions generated by CLPEX are expressed in scientific notation with four decimals. The best, worst and average values of the 10 runs of the hybrid GAs are expressed in the form of deviations from the corresponding optimal solutions generated by CPLEX (in %) according to Eq.5-2. We record the frequency reaching the optimum among the 10 runs and also calculate the CV of the 10 best solutions.

$$dev = \frac{F - F^{\mathrm{opt}}}{F^{opt}} \times 100\% \tag{5-2}$$

where F^{opt} denotes the best known solution of each instance scale and F denotes the corresponding result for each instance.

Hybrid GAs has reached the optimal solutions under 7 of 16 instances. For those instances, under which the optimal solutions have not been reached, the deviations under 3 instances are less than 1%. In worst case where n=25, delta=0.5 and fixed cost 5000, the best solution among 10 runs of hybrid GAs deviates 4.86% from the optimal solution. On average the deviations of the best solutions are 1.03% from the corresponding optimal solutions. Smaller CV indicates the solutions are more stable. However, small CV does not guarantee good performance of the algorithm. It is found that in 4 out 16 instances, e.g. instance where n=10, delta=0.1 and fixed cost 50000), CV equals to 0 for the best results of the10 runs are the same but not optimal. It indicates that the hybrid GAs is easy to fall into local optimum. Moreover, it seems that the solution quality of the hybrid GAs depends on neither instance scale (between n=10 and n=25) nor hub coverage radius and hub fixed cost.

Concerning running time, CPLEX takes less running time than the hybrid GAs under 10 instances. However, the running time of CPLEX increases rapidly with the instance scale, nearly 25 seconds for instances with 25 nodes. On the other side, none of the instances by hybrid GAs reaches 2 seconds. Moreover, no strong correlation can be distinguished between instance scale and running time of hybrid GAs. One reason is that although the string of the chromosome is shorter under small-scale instance, the crossover and mutation operations are almost the same as those for large-scale instances. Another reason is that we impose the constraint of at least 50 generations on hybrid GAs.

Generally speaking, CLPEX (exact method) outperforms the proposed hybrid GAs under small-scale instances both in running time and solution quality. However, with the increase of the instance scale, our hybrid GAs can get relatively good solutions in a significantly reduced amount of time compared with exact methods that might need exponential computation time in the worst case[413].

[413] See Blum/Roli (2003), p.269.

5.2. Tests for large-scale instances

5.2.1. Modification of input data

We choose Australia Post (AP) data set[414] for tests on large-scale instances. AP data set is first used by Ernst and Krishnamoorthy[415]. It is based on postal delivery in Sydney, Australia and consists of 200 nodes representing postal districts. Compared with other two commonly used test data sets for HLPs (CAB and Turkish data set), it has much larger demand node set with dissymmetrical flow matrix. Another feature of this data is that it displays a central business district located in the north, which means there is a large volume flowing into and out of the nodes in this district.

AP data set contains information about node coordinates, demand flow, costs for collection, transfer and distribution and hub fixed costs. We make some modifications to the data set for the computational purpose.

- **Potential hub set H and corresponding fixed cost**

In the AP data set, nodes are listed with increasing ordinate values. In order to test instance, e.g., (*H=50, N=100)*, we cannot simply take the first 50 nodes as potential hub set H, since they aggregate in the south part of the network. It is intuitively believed that nodes with large in-and-out demand flow are good locations for hubs. In this respect we take the following steps to determine the potential hub sets for different test instances. Firstly, we calculate the total demand flow originating at and destined to every demand node i with Eq.5-3.

$$f^i = \sum_{j \in N} (w_{ij} + w_{ji}) \tag{5-3}$$

Then we order f^i decreasingly and take the first 50, 80 and 100 nodes respectively as potential hub sets for the following tests. Although one can easily create potential hub sets by randomly choosing certain number of nodes from the whole data set, it makes the tests not repeatable and the corresponding results incomparable.

Meanwhile, AP data set includes two data sets, i.e. "tight" and "loose", of hub fixed cost for only 50 nodes, which is not enough for our test instances. We assign the fixed cost for nodes in potential hub set with "loose" fixed cost data set provided in AP data repeatedly. The first 100 nodes with the largest in-and-out demand flow and corresponding "loose" fixed cost values are listed in App.1, from which potential hub sets are taken for the following computational tests.

- **Scaled data in AP data set**

AP data set contains no measure unit for all data. We scale them up or down and assign unit to them (see Tal.5-3), so that the scale among them is more reasonable for our test instances. The symbols with superscript "star" denote original AP data, while the corresponding symbols without "star" denote the data in our test instances.

[414] The data is available from the OR-Library http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/phub1.txt.

[415] See Ernst/ Krishnamoorthy (1996), pp.139-154.

Symbol	Description	Unit	Conversion
w_{ij}	Daily demand	ton	$w_{ij} = w_{ij}^* \times 100$
f_k	Hub fixed cost	USD	$f_p = f_p^* \times 100$
d_{ij}	Distance	kilometer	$d_{ij} = d_{ij}^* / 100$

Table 5-3: Scaled data in the AP data set

- **Cost for collection, transfer and distribution**

The collection, transfer, and distribution cost rates in original AP data set are 3, 0.75 and 2, respectively. However, such cost rates combination is unreasonable for air-ground transportation system. We set ground and air cost rate as 0.7 (per ton*kilometer) and 0.5 respectively, just as in our real-life instances.

- **Hub coverage radius**

This is a special parameter for our covering problem. After we have scaled and adjusted the original AP data, we set hub coverage as 300 kilometers.

5.2.2. Parameter setting for hybrid GAs

How to choose control parameters for heuristics and meta-heuristics has been studied in both analytical and empirical researches. Appropriate parameters are critical for efficient algorithms. Determination of those interactive parameters itself can be a tough task[416].

We set control parameters roughly and even somewhat subjectively for tests on the performance of hybrid GAs under small-scale instances. However, the performance of GAs under large-scale instances is largely affected by control parameters. We can regard the job of GAs as a combination of exploration of new promising regions in search space and exploitation of already sampled regions. The balance of the combination, which determines the performance of GAs, is enhanced by the right choice of those important control parameters. In this section, we make preliminary tests to set control parameters for GAs under large-scale instances.

Former experiments indicate that no particular value for a parameter can be determined as the best for all instances[417]. We try to find compromise algorithm parameters that yield maybe not the best but relatively good solutions under all concerned test instances. All the tests in this section are based on SGAs that includes none of the improvement techniques proposed in Sec.4.2. In other words, the solutions are initially randomly generated and updated in every generation without injection mechanism. After hubs are determined by reproduction procedures in GAs, demand nodes are allocated to "home" hub according to "maximum flow pattern" and a PLS is followed directly afterwards. No LS on hub location decision or demand allocation decision is implemented on the final solution pool of the SGAs. Three instance scales, i.e. $(H=50, N=200)$, $(H=80, N=200)$ and $(H=100, N=200)$, are considered in each test group.

[416] Schäffer et al. conducted a factorial experiment and analyzed variance to identify and quantify the influence of the control parameters, including population size, crossover probability, and mutation probability on the performance of genetic search. See Schäffer et al (1989), pp. 51-60.

[417] See Resende/Ribeiro (2002), pp. 219-249.

- **Parameters for reproduction**

Population size N_{pop}, Crossover probability P_{cro} and Mutation probability P_{mut} are three interactive parameters for reproduction procedure in GAs.

Population size says how many chromosomes or individuals are in one generation. On the one hand, when N_{pop} is small, GAs run fast by searching in a relatively small solution space. But it results in lower population diversification; hence the algorithms converge fast to local optimum, which is also called premature. On the other hand, if there are too many individuals in the solution pool, GAs presents somewhat randomness and hesitates to converge. Former studies indicate that the best population size depends on the encoding method and the length of encoded string. It also interacts with crossover probability and mutation probability[418].

Crossover probability determines how much percent of parents will undergo crossover operation. Crossover is implemented in the hope that offspring can combine good parts from their parents and can be better than their parents. It is the primary operator in GAs to generate new individuals. In this ground, Crossover probability P_{cro} is usually high.

Mutation probability determines how much percent of individuals are mutated. Mutation serves as a secondary operator to explore new search region by altering bits of gene randomly. It is also an important mechanism to prevent GAs from falling into local extreme. "Large" Mutation probability increases diversification of the solution pool during the process of GAs, but may destroy building blocks in individuals thus disturb convergence of the algorithm. "Too large" Mutation probability tends to drive the GAs to search randomly[419], while "too small" Mutation probability may result in premature of the algorithm.

Although the choice of control parameters is case by case, researchers have proposed two combination strategies that show relatively good performance. One has large Population size but small Crossover and Mutation probabilities, while the other has small Population size but large Mutation and Crossover probabilities[420]. Both of the two combinations of parameters setting can maintain the dynamic balance between diversification and intensification of the solution pool.

As it is quite time consuming to find feasible solutions in our case, we take up the second strategy by creating a relatively small solution pool with large Crossover and Mutation probabilities. We fix Population size N_{pop} at 50 subjectively. We consider three instance scales, i.e., $(H{=}50, \; N = 200)$, $(H = 80, N = 200)$ and $(H = 100, N = 200)$. For each instance scale, we take a combination of P_{cro} with variations of 0.5, 0.7 and 0.9 and P_{mut} with variations of 0.01, 0.1, 0.2 and 0.3. We run the preliminary tests 10 times for each instance. The programming runs at least 50 generations and at most 200 generations. It also terminates when the average objective value of individuals in the solution pool has less than 0.001% improvement after 5 generations. Tab.17 summarizes the results of 10 runs for each instance. The best solutions under each instance scale, i.e. $(H{=}50, \; N = 200)$, $(H = 80, N = 200)$ and $(H = 100, N = 200)$, we have found are 62743272, 61017632 and

[418] See DeJong/ Spears (1991), pp.38-47.

[419] See Xu et al (2006), p.600.

[420] See DeJong/ Spears (1991), pp.38-47; Grefenstette (1986), pp.122-128.

60622912, respectively. To make it easy to read, we fill the table with the deviation of corresponding values from the best known solution F^* of each instance scale (calculated with Eq.5-4). The column headings list Mutation probability for different problem scales and the line headings illustrate "deviation of best solution in solution pool" (in %), "deviation of average value of solutions in the solution pool" (in %) and CV of 10 best solutions for each Crossover probability.

$$dev = \frac{F - F^*}{F^*} \times 100\% \tag{5-4}$$

where F^* denotes the best known solution of each instance scale and F denotes the corresponding result for each instance.

			Crossover probability								
			0.5			0.7			0.9		
			B. dev.	**A. dev.**	**CV(best)**	**B. dev.**	**A. dev.**	**CV(best)**	**B. dev.**	**A. dev.**	**CV(best)**
Mutation probability	**H=50**	0.01	0.87%	13.39%	0.0003	1.33%	12.77%	0.0131	0.29%	11.00%	0.0046
		0.1	0.68%	12.46%	0.0046	2.25%	13.67%	0.0153	0.64%	11.95%	0.0046
		0.2	0.33%	10.41%	0.0105	0.00%	10.30%	0.0000	0.10%	10.49%	0.0030
		0.3	0.01%	9.47%	0.0001	0.02%	9.42%	0.0030	0.67%	11.49%	0.0298
	H=80	0.01	1.47%	9.81%	0.0013	0.00%	9.36%	0.0000	1.25%	9.00%	0.0146
		0.1	0.01%	8.64%	0.0002	0.88%	8.67%	0.0141	0.29%	9.09%	0.0093
		0.2	0.59%	8.07%	0.0123	0.00%	8.26%	0.0000	0.00%	9.89%	0.0000
		0.3	0.00%	8.22%	0.0000	0.76%	8.26%	0.0132	0.00%	9.34%	0.0000
	H=100	0.01	3.60%	10.49%	0.0265	3.55%	10.89%	0.0238	2.01%	9.95%	0.0000
		0.1	0.35%	7.28%	0.0051	1.46%	8.36%	0.0119	1.65%	8.00%	0.0240
		0.2	0.11%	8.09%	0.0037	0.09%	7.00%	0.0030	1.06%	6.88%	0.0114
		0.3	0.20%	8.29%	0.0035	2.71%	9.94%	0.0171	1.02%	10.16%	0.0135

Table 5-4: Deviation from the best-known solution under each instance scale (tests for reproduction parameters)

A. dev.: average deviation of 10 runs of the average solution in the final solution pool from the best-known solution of that instance scale

B. dev.: average deviation of 10 runs of the best solution in final solution pool from the best-known solution of that instance scale

CV (best): CV values of the best solutions of 10 runs

After checking the results of all the combinations of (P_{cro}, P_{mut}) under different instance scales (see Tab.5-4), we empirically conclude that when the Crossover probability P_{cro} and Mutation probability P_{mut} is 0.7 and 0.2 respectively, GAs can yield relatively good solutions under pertinent instance scales. Comparatively speaking, it is an ideal parameter combination, although it is not the best for all instances. Particularly, when H=50, (0.7, 0.2) performs the best in both "B. dev." and "A. dev.", although (0.5, 0.3) and (0.7, 0.3) also yield almost the same "best solution"; when H=80, the "B. dev." of (0.5, 0.3) and (0.9, 0.2) is the same as (0.7, 0.2); when H=100, the "A. dev." of (0.7,0.2) is larger than that of (0.9, 0.2) and it generate the best "B. dev.".

When we consider the index of CV, the results of instances with (P_{cro}, P_{mut}) as (0.7, 0.2) under different scales are quite stable. However, low values of CV do not guarantee good performance of the algorithm. The CV values of instances $(H=50, P_{cro}=0.5, P_{mut}=0.01)$ and $(H=80, P_{cro}=0.5, P_{mut}=0.01)$ are relatively low, but the solution qualities are not good. This may indicate that the algorithm tends to be premature with small crossover and mutation probability.

We make T-tests with SAS[421] on both "B. dev." and "A. dev." between (0.7, 0.2) and other parameter combinations under different instance scales to support this selection. The test results show that the differences are not always significant, taking examples of the best and average solutions between $(H=50, P_{cro}=0.7, P_{mut}=0.2)$ and $(H=50, P_{cro}=0.9, P_{mut}=0.2)$ (see App.2).

With regards to these reasons, reproduction parameters combination of (0.7, 0.2) can stably yield good solutions under different instance scales, although it is not significantly better than all the other parameter combinations under all pertinent instance scales. Hence, a Crossover probability of 0.7 and a Mutation probability of 0.2 are hereafter used in the later tests for GAs.

Generally speaking, the solutions under different instance scales conform to the trend that the total cost of the network decreases with the increasing size of the potential hub set H. However, there is also anomaly, which is reflected by, e.g. $(H=100, P_{cro}=0.5, P_{mut}=0.01)$. It probably results from the premature of the algorithms. As we can observe in Tab.18, those instances with "bad" solutions, compared with others under the same instance scale, terminate more often than not at earlier generation. It can be inferred that the corresponding reproduction parameters are irrational. Since we force the programming to run at least 50 generations, we can anticipate that those instances with generations only few over 50 are premature in most runs.

Generation			**Crossover probability**		
			0.5	**0.7**	**0.9**
Mutation probability	**H=50**	0.01	51.3	53.0	54.4
		0.1	67.5	57.8	64.0
		0.2	61.8	60.7	60.9
		0.3	65.2	62.5	63.5
	H=80	0.01	52.0	63.0	52.9
		0.1	58.0	62.8	53.6

[421] T-test, more specifically speaking, independent two-sample t-test is used to test whether two population means are different.

		0.2	56.4	59.4	53.8
		0.3	55.7	60.9	55.6
	H=100	0.01	53.4	52.5	51.8
		0.1	73.7	53.0	59.3
		0.2	64.6	61.9	63.0
		0.3	72.0	63.9	69.5

Table 5-5: Average generation number under each instance

At the first glance of Tab.5-5, we cannot find obvious correlation between generation number and Mutation probability, Crossover probability or instance scale. After we make summaries according to instance scale (see Tab.5-6), Crossover probability (see Tab.5-7) and Mutation probability (see Tab.5-8) respectively, we find some interesting results.

H	**50**	**80**	**100**
Generations	60.22	57.01	61.55

Table 5-6: Average generation number according to instance scale

Crossover probability	**0.5**	**0.7**	**0.9**
Generations	60.97	59.28	58.53

Table 5-7: Average generation number according to Crossover probability

Mutation rate	**0.01**	**0.1**	**0.2**	**0.3**
Generations	53.70	61.11	60.29	63.21

Table 5-8: Average generation number according to Mutation probability

The generation number does not increase with instance scale (see Tab.5-6), which seems to be a little surprising at first. It was supposed to be difficult for GAs to converge when search space is large. However, the feature of our problem may be the reason of this phenomenon. Once a node is chosen as hub, it can cover many other potential hubs, dramatically narrowing the search space for GAs. Consequently, the computational time does not dramatically increase with the size of potential hub set *H*.

We also find that the generation number is roughly inversely proportional to Crossover probability (see Tab.5-7). It may be explained that large Crossover probability intensifies the search strength and accelerates the convergence of the algorithms.

On the contrary, generation number roughly increases with Mutation probability, with a small deviation from this trend between the instances of $P_{mut} = 0.1$ and $P_{mut} = 0.2$ (see Tab.5-8). On the one hand, "large" Mutation probability increases the diversification of solution pool and broadens the search space of the algorithm, hence increasing the odds of finding the global optimum. However, the impact of large Mutation probability can be slow convergence of the algorithm and too large Mutation probability may lead the algorithms to run randomly. On the other hand, "small" Mutation probability is liable to be the cause of prematurity. As we can observe, the phenomenon of prematurity is exacerbated under instances with $P_{mut} = 0.01$, when most runs stop at 50th generation, which is the minimum generation we impose on the algorithm. We can therefore infer that 0.01 is a too small Mutation probability value. More rational Mutation probability is a medium one that can not only bring divergence to the solution pool but also prevent early convergence of the algorithm.

When we analyze Tab.5-4 and 5-5 together, we cannot find positive correlation between generation number and solution quality, especially in terms of "average values of solutions in the final solution pool". In other words, if the algorithm convergences at an early stage (premature), the result is always not so good, for instance $(H{=}100，P_{cro} = 0.5, P_{mut} = 0.01)$, $(H{=}100，P_{cro} = 0.7, P_{mut} = 0.01)$ and $(H{=}100，P_{cro} = 0.9, P_{mut} = 0.01)$. However, the solutions are not guaranteed to be good when generation number is large, for instance $(H{=}100，P_{cro} = 0.7, P_{mut} = 0.1)$ and $(H{=}100，P_{cro} = 0.7, P_{mut} = 0.2)$.

Running time (s)			Crossover probability		
			0.5	**0.7**	**0.9**
Mutation probability	**H=50**	0.01	283.0/5.52	360.7/6.79	523.7/9.63
		0.1	359.3/5.32	398.3/6.89	575.6/8.98
		0.2	398.9/6.44	467.7/7.71	596.8/9.79
		0.3	426.5/6.54	500.1/8.00	616.7/9.71
	H=80	0.01	309.7/5.96	403.7/6.41	414.3/7.83
		0.1	337.3/5.82	429.7/6.84	419.7/7.83
		0.2	336.7/5.97	468.1/7.88	429.4/7.97
		0.3	330.9/5.94	477.5/7.84	444.2/7.99
	H=100	0.01	237.4/4.44	279.7/5.31	356.3/6.88
		0.1	342.3/4.64	301.0/5.68	458.3/7.73
		0.2	323.0/5.00	360.5/5.82	496.7/7.88
		0.3	368.5/5.12	367.2/5.75	567.6/8.17

Table 5-9: Total running time & running time per generation

Tab.5-9 displays the "total running time" and "running time per generation" with separation by slash. The later one is calculated by dividing the "total running time" with the corresponding "generation number". Higher Crossover and Mutation probability are likely to entail more computational time, which is easy to understand.

We also compare the running time for large-scale instances in Tab.5-9 with that of small-scale instances in Tab. 5-2. It grows mild with the instance scale, from less than 2 seconds for N=25 to about 300-600 seconds for N=200. However, CPLEX takes less than 0.1 second for N=10, but nearly 25 seconds for N=25. In other words, the running time of exact algorithms grows exponentially with the instance scale, while the running time for proposed hybrid GAs experiences a linear increase with the instance scale. This is the major motivation for us to adopt GAs for our real-life large-scale instances.

- Parameters for termination

We try to set a value for parameter "Maximum generation $N_{\max\_gen}$" to guarantee the algorithm has fallen into local optimum. Meanwhile, we timely stop the algorithm with the parameter "Maximum number of iteration without improvement N_{imp}" to save computational resource. Moreover, we force the algorithms to run at least $N_{\min\_gen}$ number of generations to prevent the prematurity of the algorithms.

In preliminary tests for setting reproduction parameters P_{cro} and P_{mut} , we set the termination criterion like this: the programming runs at least 50 generations and at most 200 generations. It also terminates when the average value of solutions in the solution pool after 5 generations makes less than 0.001% improvement. As we can observe from Tab.18, the average generation of all instances is about 60. Among all the 360 runs of the algorithms, there are only 1 case with more than 100 generations, 1 between 90-100 generations, and 3 between 80-90 generations. So this termination criterion allows GAs to converge to local optimum under all the instances we consider.

Fig.5-2 and 5-3 record "average" and "best" solution values in the solution pool during one implementation of GAs under the instances of $(H{=}50，P_{cro}=0.7, P_{mut}=0.2)$, $(H{=}80，P_{cro}=0.7, P_{mut}=0.2)$ and $(H{=}100，P_{cro}=0.7, P_{mut}=0.2)$ respectively. The X-coordinate represents "number of generations". The "average value of solutions in the solution pool" in Fig.38 shows the convergence of the algorithms. The test results show that the solution pool converges fast before 30th generation. The speed of convergence slows down afterwards. The programming terminates before 70th generation under all the three instances. In other words, individuals in the solution pool can't be improved even the programming is prolonged. Furthermore, the best solution in final solution pool occurs at about 30th generation. It can be hardly improved with further iteration. We believe it has reached the local optimum.

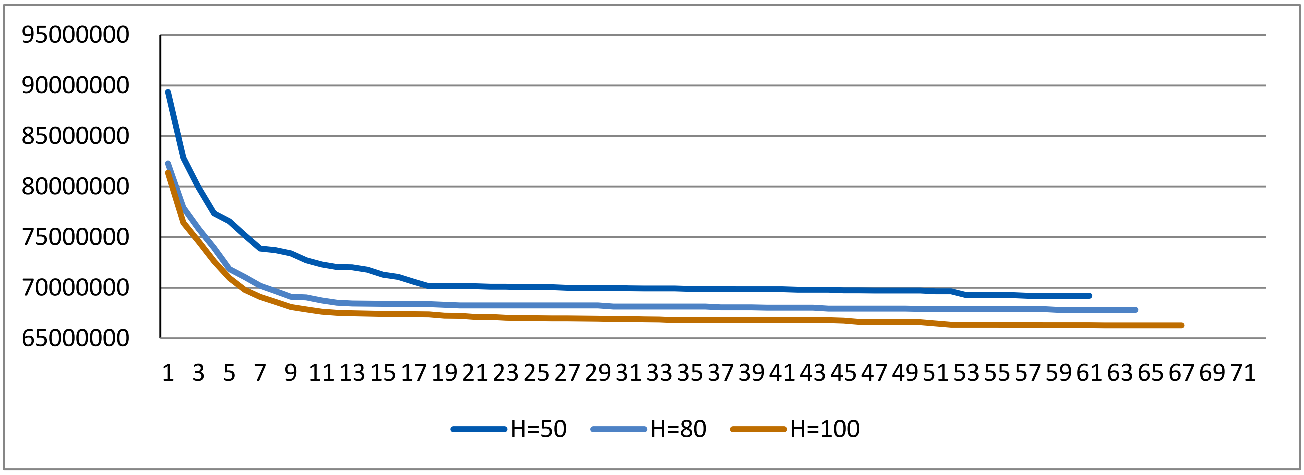

Figure 5-1: Average value of solutions in the solution pool(with Crossover and Mutation probability as 0.7 and 0.2)

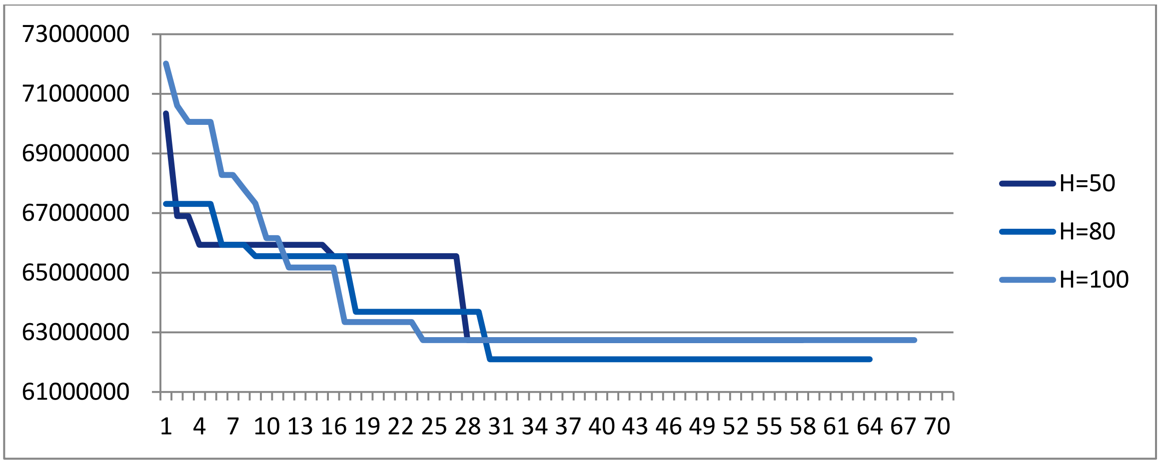

Figure 5-2: Best solutions in the solution pool (with Crossover and Mutation probability as 0.7 and 0.2)

5.2.3. Tests for improvement techniques

In this section, we test the performance of the first four improvement techniques proposed in Sec.4.2 based on basic model, leaving the last one in the next chapter, which must be tested on extension models. All the tests are based on data set $(N = 200, H = 50)$.

Concerning the instance scale of $(N = 200, H = 50)$, we have more than 2 billion binary variables and 1 billion linear constraints with the basic model, which are beyond the capacity of CPLEX[422] and also our ability to solve it optimally with other customized algorithm. For this reason, we evaluate the performance of the proposed improvement techniques under large-scale instances by comparing the corresponding solutions with the solutions by the untailored SGAs, since the optimal solutions are not available.

5.2.3.1 Tests on constructive procedure

In this section we test if constructive procedure proposed in Sec.4.2.1 can improve the performance of GAs by producing "good initial solutions".

It is anticipated that a "good" parameter α is dependent on the solution space of the instance[423]. So we consider two instances, with hub coverage radius as 300 kilometers and 200 kilometers, respectively. We denote them as Hub (300) and Hub (200) in the following. The solution of the latter case contains more hubs, which denotes larger solution space.

It can be expected that CGA (100%)[424] yields the same initial solution each iteration if the first element is also selected from RCL as other cases. In order to generate enough solutions for the initial solution pool, we traverse through set H rather than take the top node in RCL as the first node.

[422] IBM ILOGCPLEX Optimizer has solved problems with millions of constraints and variables. See http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/ (access on 20.01.2013).

[423] See Delorme (2004), p.566.

[424] We denote CGAs with parameter α as CGAs (α) in the following.

We test the influence of parameter α to the performance of GAs under instance $(N=200, H=50)$. We set α from 0 to 100% with a step length of 25%. In SGAs most of the randomly generated initial solutions are infeasible. We just delete the infeasible solutions until enough feasible initial solutions are generated. We keep the other parameters for all the instances the same as follows: Population size N_{pop} 50, Maximum generation $N_{\max\_gen}$ 200, Minimum generation $N_{\min\_gen}$ 50, Maximum number of iteration without 0.001% improvement N_{imp} 5, Crossover probability P_{cro} 70%, and Mutation probability P_{mut} 20%.

We test the performance of the proposed algorithms with the hub radius as 200 and 300 respectively. We also compare CGAs with SGAs. There're 6 instances under each hub radius, with 10 independent replications for each instance. As a matter of fact, we first test the instance of Hub (300), whose results are somewhat disappointing: all the 10 best solutions designate the same 3 nodes as hubs. We believe that it is the limited search space leads to the inefficacy of our improvement technique. Then we test the instance of Hub (200), whose results display clear tendency and strongly support our idea. The best solutions of Hub (200) are with 5 or 6 hubs. In other words, Hub (200) has larger search space for GAs than Hub (300) so that there is larger improvement space for GAs under Hub (200) than under Hub (300). It may be because of this reason that GAs can take advantage of its strengths more effectively under instance Hub (200).

In Fig.5-4 we plot the running time (in second) with different hub radii and values of α to generate initial solutions by constructive procedure. We can observe that it is more time consuming for instances with larger randomness, i.e. larger search space. Specifically, when α decreases from 100% to 0, the constructive procedure turns from greedy one into random one. Moreover, for a fixed choice of α, feasible solution space becomes larger as hub radius decreased. In both cases, the algorithms become more time consuming.

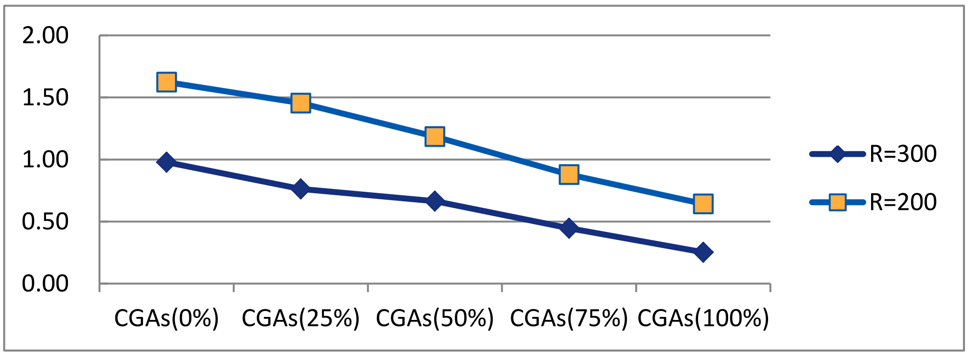

Figure 5-3: Running time (in second) to generate initial solutions with constructive procedure

First of all, we analyze the results of tests under Hub (200). The best-known result of Hub (200) we have found is 73391458. Tab.5-10 reports the deviation of the initial best solution & final best solution from the best-known solution calculated across the 10 runs with Eq.5.4. Tab.5-10 also reports the CV values of the 10best initial and finial solutions under each instance.

	Initial solution			Final solution			Imp on Avg.
	B. dev.	CV(best)	A. dev.	B. dev.	CV(best)	A. dev.	
SGAs	68.48%	0.1737	91.57%	0.87%	0.0129	6.65%	44.33%
CGAs(0%)	40.12%	0.0519	85.03%	0.67%	0.0062	6.05%	42.68%
CGAs(25%)	28.66%	0.0673	82.38%	0.56%	0.0098	5.96%	41.90%
CGAs(50%)	18.87%	0.0427	78.40%	1.32%	0.0079	7.39%	39.81%
CGAs(75%)	18.62%	0.0490	67.09%	1.98%	0.0256	7.83%	35.47%
CGAs(100%)	8.08%	0.0146	42.98%	0.20%	0.0018	5.53%	26.19%
Average of all	30.47%	0.0665	74.57%	0.93%	0.0107	6.57%	38.40%
Average without SGAs	22.87%	0.0451	71.18%	0.95%	0.0103	6.55%	37.21%

Table 5-10: Deviation of initial solution & final solution from the best-known solution (Hub (200))

A. dev.: average deviation of 10 runs of the average solution in the solution pool from the best-known solution of that instance scale

B. dev.: average deviation of 10 runs of the best solution in final solution pool from the best-known solution of that instance scale

CV (best): CV values of the best solutions of 10 runs

Imp on Avg.: the improvement on the average solutions in the solution pool by GAs (see Eq.5-5)

CGAs (100%) perform the best on both "best" solution and "average solution" in the final solution pool. Meanwhile, it also starts from the best initial solution pool. It strongly supports the discipline that "good initial solution" can improve the performance of GAs[425]. Moreover, it is the most stable one with the least CV values for both initial and final best results. It is believed that CGAs (100%) performances the best. This conclusion is also verified by the results of *T*-tests between CGAs (100%) and all the other instances.

The tendency of the results is also obvious that the initial solutions are getting better with increasing α in terms of best and average solution in initial solution pools. A higher α means more greediness, which is also reflected by the CV values. Specifically, the greedy algorithm constructs initial solutions of the best quality and least CV when α=100%, followed by those in the middle, and then by the random algorithm when α=0%. The randomly generated initial solutions of SGAs are of the worst quality and largest CV value among all the instances.

However, the initial solutions obtained by the constructive procedure are still far from the best-known solutions (22.87% on average and up to 40.12% larger than the best-known solution). The GAs phase improves these solutions significantly, 37.21% on average. Actually, iteration of GAs works with different strength on different initial solution pools. To analysis this effect, we calculate the improvement of average solution in the pool from initial to final one, with the table head "Imp on Avg." (see Eq.5-5).

$$Improvement\ in\ avg. = \frac{Initial\ avg. - Final\ avg.}{Initial\ avg.} \times 100\% \tag{5-5}$$

[425] See e.g.Talbi (2002), p.545.

where “Initial avg.” denotes the average solution in the initial solution pool and “Final avg.” denotes the average solution in the final solution pool.

The tendency can be easily found that the improvement from initial solution pool to final one is higher for worse initial solutions, since the difference of solution quality among final solution pools for pertinent instances is much smaller than that among initial solution pools. For example, CGAs (0%) is improved by 42.68% and SGAs is improved by 44.33%, while CGAs (100%) is only improved by 26.10%. For this reason, GAs are quite good at searching good solution in a wide space.

	Initial solution			**Final solution**			**Imp on**
	B. dev.	**CV(best)**	**A. dev.**	**B. dev.**	**CV(best)**	**A. dev.**	**Avg.**
SGAs	30.83%	0.1858	79.36%	0.19%	0.0181	10.30%	38.51%
CGAs(0%)	9.51%	0.0287	44.32%	0.19%	0.0040	10.32%	23.56%
CGAs(25%)	11.92%	0.0240	46.00%	0.38%	0.0049	11.37%	23.72%
CGAs(50%)	9.55%	0.0218	44.80%	0.10%	0.0030	10.17%	23.92%
CGAs(75%)	7.87%	0.0031	40.61%	0.10%	0.0030	9.17%	22.36%
CGAs(100%)	18.70%	0.0478	73.63%	0.19%	0.0040	10.31%	36.47%
Average	14.73%	0.0519	54.79%	0.16%	0.0062	10.27%	28.09%

Table 5-11: Deviation of initial solution & final solution from the best-known solution (Hub (300))

Then we analyze the test results under Hub (300) (see Tab.5-11). The table titles have the same meaning as those in Tab.5-10. The best-known result of Hub (300) we have found is 62743272. The tendency we have mentioned in analysis for instance Hub (200) is also applicable to instance Hub (300) with few exceptions. The tendency of initial solution quantity among different instances is faint compared with that of Hub (200). The improvement function of GAs that is reflected by the index “Imp on Avg.” is also moderate under Hub (300). Moreover, most of the T-tests on final best solutions between different instances are not significant. All of these may result from the limited solution space of Hub (300), compared with that of Hub (200).

When we compare the results of CGAs with SGAs, we find that CGAs does not always work positively on the performance of the algorithms (see Tab.5-12 and 5-13). The values listed in the two tables, i.e., the improvement of best solution and average solution in the final solution pool by CGAs based on SGAs, are calculated with Eq.5-6. Take Hub (200) as an example, the instances CGAs (50%) and CGAs (75%) even have worse solutions than SGAs. For Hub (300), SGAs have a similar performance as CGAs (0%) and CGAs (100%), while it has a better performance than CGAs (25%). So a rational parameter α is vital for the algorithm performance.

$$improvement = \frac{F_{SGA} - F_{CGA}}{F_{SGA}} \times 100\% \tag{5-6}$$

where F_{SGA} denotes the best or average result in the final solution pool of SGAs, while F_{CGA} denotes the counterparts of CGAs.

Hub (200)	Best solution	Average in solution pool
SGAs	0.00%	0.00%
CGAs(0%)	0.19%	0.56%
CGAs(25%)	0.30%	0.64%
CGAs(50%)	-0.45%	-0.69%
CGAs(75%)	-1.11%	-1.11%
CGAs(100%)	0.86%	1.05%

Table 5-12: Improvement of CGAs on SGAs (Hub (200))

Hub (300)	Best solution	Average in solution pool
SGAs	0.00%	0.00%
CGAs(0%)	0.00%	-0.02%
CGAs(25%)	-0.19%	-0.98%
CGAs(50%)	0.09%	0.12%
CGAs(75%)	0.09%	1.02%
CGAs(100%)	0.00%	-0.01%

Table 5-13: Improvement of CGAs on SGAs (Hub (300))

However, half of the CGAs instances generate improved results in terms of both best and average solutions, and 80% of the instances improve in either one. Moreover, although SGAs have the potential to generate as good solutions as CGAs, its performance is not stable since it always has the largest CV values under all instances. Thus the solution of CGAs is more stable than that of SGAs in this sense.

Although the running time of CGAs is longer than SGAs under all instances, the time is worthy for better solutions. Better solutions from CGAs imply that constructive procedure works well to enhance the performance of GAs by finding "good" initial solutions. Observe that, the randomness in constructive procedure brings diversity in the initial solution pool, while the greedy function is to generate "good" solutions. High quality solutions as well as large solution diversity are desirable characteristics of initial solutions for GAs[426].

5.2.3.2 Tests on injection mechanism

Injection mechanism increases the diversity of the solution pools during the process of GAs, while constructive procedure adds the diversity to the initial solution pool. It is anticipated that these two measures have interrelated influence on the final solution quality, i.e. both of them act on the balance of diversification and intensification of the solution pool of GAs. In this respect, we make computational tests with different combinations of Injection rate P_{new} and constructive parameter α . Specifically, to distinguish the impact of injection intensity on the performance of GAs, we take P_{new} as 0%, 4%, 8% and 12%. Each injection rate is considered together with 6 different initial solution generation procedures, namely those in SGAs, CGAs (0%), CGAs (25%), CGAs (50%), CGAs (75%) and CGAs (100%). There are totally 24 instances, 10 tests for each instance.

[426] See e.g. Hertz/ Kobler (2000), pp.1-12.

For all these tests[427], we eliminate the termination criterion "at least 50 generations" to check the convergence of algorithms without disturbance. We use the same parameters as those in SGAs, that is, Population size 50, Maximum generations 200, Maximum number of iteration without 0.001% improvement 5, Crossover probability 70%, Mutation probability 20%, Hub radius 300 and "maximum flow allocation" pattern. It is believed that injection mechanism may impact on the best choice of Crossover and Mutation probability. However, it is beyond our research to find the best parameter combination of Crossover probability, Mutation probability and Injection rate, or even together with Population size. Our intent here is to check if injection mechanism can further enhance the performance of GAs. Tab.5-14 summarizes the results of 240 tests expressed in deviation from the best-known solution (62743272) for injection mechanism. The title "Gen." denotes the average generation number of the 10 runs under each instance, while others have the same meaning as those in Tab.5-10.

[427] The instances under Injection 0% are just those for constructive procedure. We do these tests again here but without the termination criterion "at least 50 generations". The resultant solutions are quite similar as before.

Initial solution	Injection 0%				Injection 4%				Injection8%				Injection 12%			
	Gen.	B. dev.	CV (best)	A. dev.	Gen.	B. dev.	CV (best)	A. dev.	Gen.	B. dev.	CV (best)	A. dev.	Gen.	B. dev.	CV (best)	A. dev.
Constructive(0%)	31.5	0.19%	0.0040	10.78%	31.2	0.00%	0.0000	8.37%	34.6	0.00%	0.0000	8.17%	39.0	0.19%	0.0040	11.30%
Constructive(25%)	32.3	0.38%	0.0049	11.01%	30.4	0.00%	0.0000	8.26%	31.8	0.00%	0.0000	8.17%	32.8	0.00%	0.0000	8.22%
Constructive(50%)	39.8	0.19%	0.0040	10.17%	27.8	0.00%	0.0000	8.20%	22.8	0.00%	0.0000	8.12%	38.5	0.19%	0.0040	10.15%
Constructive(75%)	40.4	0.10%	0.0030	9.36%	32.2	0.00%	0.0000	7.83%	32.9	0.00%	0.0000	7.83%	31.2	0.00%	0.0000	7.93%
Constructive(100%)	36.5	0.19%	0.0040	10.99%	47.4	0.00%	0.0000	8.98%	41.4	0.00%	0.0000	9.35%	35.0	0.00%	0.0000	9.66%
SGAs	36.1	0.21%	0.0040	10.46%	31.8	0.58%	0.0049	9.42%	28.0	0.00%	0.0000	8.38%	41.2	0.19%	0.0040	8.46%

Table 5-14: Test results of injection mechanism

The test results show that the introducing the injection operator into CGAs (see the first 5 rows) can bring much better solutions than solutions by CGAs without injection. In particular, CGAs with injection find the best-known solution for all 10 runs under all instances except the instance [Constructive (0%), Injection 12%]. Moreover, all these average solutions in final solution pools are better than those without injection. For each initial solution generation procedure (each row), we test the significance between instance of Injection 0% and instance of Injection 4%, 8% and 12% in terms of both best and average solution in the final solution pool. All the *T*-test results support the conclusion except the results in terms of average solution relating instances [Constructive (0%), Injection 12%].

However, the test results actually do not indicate which Injection rate is better for our instance. Although the results in Tab.5-14 show that the average solution in final solution pool first decrease then increases with the increasing Injection rate, the results by *T*-tests indicate that the difference between Injection 4% and 8% is not significant under all instances in terms of both best and average solution. We may only conclude that a moderate Injection rate is preferred by GAs. Injection operator brings diversity to the solution pool, for the case where the injection operator is not used the solution seems to be easier to stagnate at a local optimum. However, too high Injection rate may break up the balance between diversity and intensity, leading GAs to randomness.

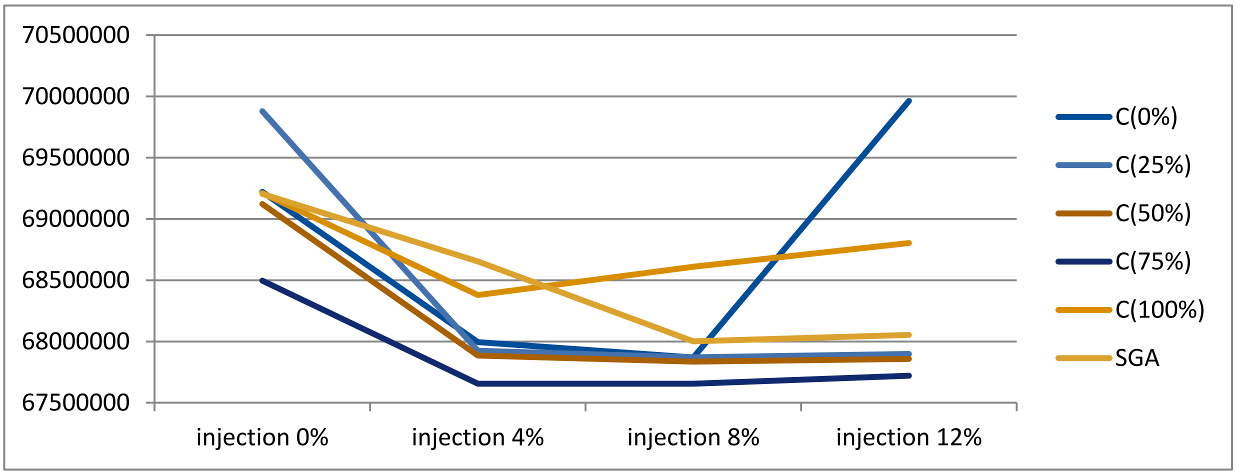

Figure 5-4: Average objective value of final solution pool with different initiation policies

The test results also show that the improvement of injection mechanism on SGAs is not as obvious as that on CGAs. This conclusion is verified by the *T*-test results between (SGAs, Injection 0%) and (SGAs, Injection 4%) and results between (SGAs, Injection 0%) and (SGAs, Injection 12%).

When Injection rate is fixed, average solution in final solution pool first decreases then increases, with the exception of instance [Constructive (25%), Injection 0%]. All instances achieve the best average solutions at Construction (75%) (see Fig.5-5), which is not totally supported by the results of *T*-tests of average solutions on some instances between Construction (75%) and corresponding instances of Construction (100%) and Construction (50%). But results of the most *T*-tests support the idea that the diversity in the solution pool during the GAs process is as important as that in the initial solution pool. Contrary to our anticipation, the effect of injection mechanism and constructive procedure is relatively independent and can hardly compensate to each other.

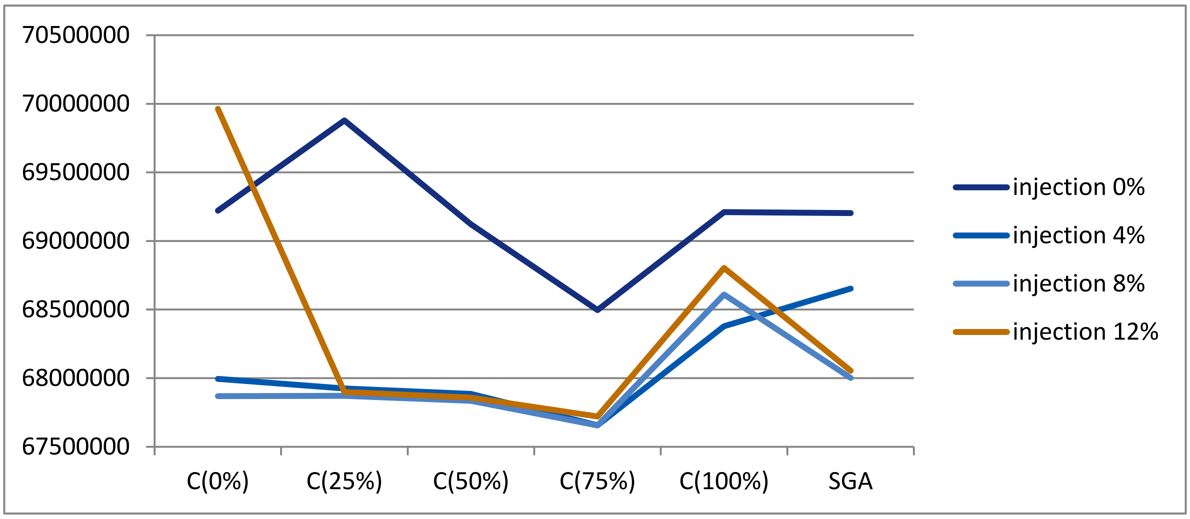

Figure 5-5: Average objective value of final solution pool with different injection policies

In terms of convergence rate, we cannot find strong correlation of generation number among different Injection rates. But under the preferred instance of Constructive (75%), the difference between instances with injection and without injection is obvious. Fig.5-6 describes typical examples of best and average solutions with Injection rate 8% and without injection, respectively. We impose the terminate criterion "at least 50 generations" to see the behavior of the algorithms more clearly. The *X*-coordinate denotes the number of generation, which can also be viewed as a representation of the computational time, while the *Y*-coordinate is the best and average solutions in solution pool. This result indicates the effectiveness of incorporating injection mechanism into GAs in terms of solution quantity and convergence rate. GAs with injection operator converges obviously faster than that without injection in both best and average solutions even they are initially similar. Although both cases find the best-known solutions, GAs without injection find the best-known solution at 51th generation, while GAs with injection at 16th generation. The average solution in final solution pool is much better for GAs with injection than that without injection. If the termination criterion "at least 50 generations" is eliminated in GAs without injection, the chance to find the best solution is much less.

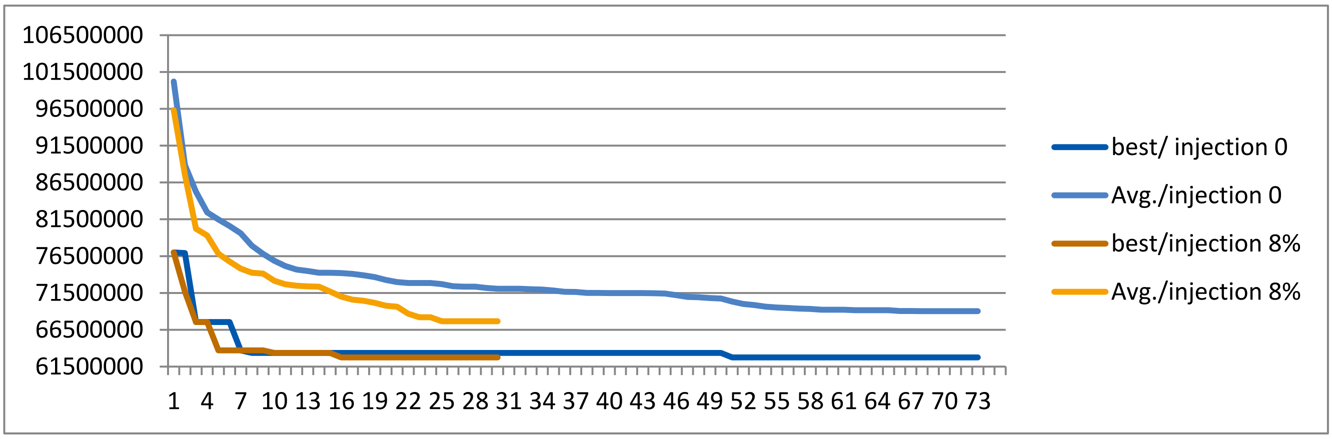

Figure 5-6: Discrepancy in solutions with and without injection mechanism (constructive 75%)

5.2.3.3 Tests on LS after GAs

In this section, we test if LS on hub location and demand allocation decisions after GAs can further improve the best solution.

Test 1: FLS on hub location and demand allocation decision

First of all, we follow the improvement techniques proposed in Sec.4.2.3. LS on hub location and demand allocation decisions is carried out fully. In particular, FLS on demand allocation decision is embedded in FLS on hub location decision until all the hub swap possibilities have been tried[428].

This test is carried out after GAs under instances of Hub (300) and Hub (200), respectively. Parameters for GAs with Hub (300) are set as follows: Constructive (75%), Population size 50, Injection 8%, Mutation probability 20% and Crossover probability 70%. Constructive parameter α is changed to 100% under instance Hub (200). These are the most efficient parameters we have found for these two instances. Tab.5-15 displays the results of tests on one final solution pool of GAs under both instances (not summary of 10 runs).

		Hub (300)	**Hub (200)**
GAs result	Average solution	64150008	75036932
	Best solution	62743272	73391458
	Selected hubs (best solution)	27 35 140	7 24 27 93 154
	Running time (s)	280.848	233.378
LS 1	Average solution	64095193	74972814
	Best solution	62689660	73263979
	Selected hubs (best solution)	27 35 140	7 24 27 93 154
	Running time (s)	113.054	122.148

Table 5-15: Test results of full local search on hub location and demand allocation decisions

The test results are somewhat surprising. After the FLS on both hub location and demand allocation decisions, the selected hubs do not change under both instances. Such results confirm the effective exploration of GAs. Although FLS has made no improvement on best hub location decision, it has made substantial improvement on other individuals in the solution pool. More specifically, 86% of the individuals in the two final solution pools have changed the hub location. Meanwhile, nearly all the individuals are improved by allocation LS. It is shown in Tab.28 allocation LS improves the best solution in final solution pool for both instances by 0.086% and 0.174%, respectively.

However, the overall FLS procedure is computationally expensive, although small improvement means huge money in reality. The FLS takes up between one fourth and one third of the total running time (28.7% and 34.4%).

[428] FLS does not equal to enumeration, since we only explore the specific neighborhood.

Test 2: FLS on demand allocation decision

The findings of Test 1 indicate that FLS may be not suitable for GAs in our case, considering computational time and improvement effect. In this respect, we apply only FLS on demand allocation decision after GAs. As we can easily expect, FLS on demand allocation decision runs much faster than FLS on hub location decision. Tab.5-16 displays the test results of FLS on demand allocation decision under instances of Hub (300) and Hub (200).

		Hub (300)	**Hub (200)**
GAs result	Average solution	64150008	75036932
	Best solution	62743272	73391458
	Selected hubs (best solution)	27 35 140	7 24 27 93 154
	Running time (s)	280.848	233.378
LS 2	Average solution	64095193	74972814
	Best solution	62689660	73263979
	Selected hubs (best solution)	27 35 140	7 24 27 93 154
	Running time (s)	2.293	2.169

Table 5-16: Test results of full local search on demand allocation decisions

The best solutions under both instances after Test 1 and 2 are the same, since the LS on hub location decision does not make any improvement. But the running time for FLS on allocation decision is much less than that on hub location decision, i.e. 0.81% and 0.92% of the total running time under the two instances respectively. The FLS on allocation decision improves nearly all the individuals in the final solution pool.

When we compare the average solutions in Test 1 and Test 2, we find that the improvement by FLS on allocation decision is tiny if hub location decision is not enhanced. In other words, the hub location decisions play a decisive role in determining the solution quality. This consolidates the idea to apply PLS on demand allocation decision during the process of GAs, since it is not worth much effort to improve the demand allocation decision until the hub location decision is good enough.

5.2.3.4 Tests on different allocation patterns before LS

In this section, we test how a starting point of LS on demand allocation decision can impact on the solution quality. We would like to find the best allocation pattern for our algorithm by comparing the solutions of hybrid GAs with different allocation patterns, i.e. "the nearest-distance", "multi-criterion", "maximum-flow" and "minimum cost" allocation patterns. The tests are implemented under H (300) and $(N = 200, N = 50)$.We use the same parameters as those in tests for CGAs and injection mechanism: Population Size N_{pop} 50, Maximum generation $N_{\max\_gen}$ 200, Minimum generation $N_{\min\_gen}$ 50, Maximum number of iteration without 0.001% improvement N_{imp} 5, Crossover probability P_{cro} 70%, and Mutation probability P_{mut} 20%. For SGAs, we do not adopt constructive procedure for initial solution generation, injection mechanism or LS after GAs. But CGAs (75%) is conducted with Injection 8% and followed with FLS on demand allocation decision. We summarize the results of 10 runs under each instance in Tab.5-17 based on the updated best-known solution (62689660) by the tests on "LS after GAs".

	Allocation pattern	Running time(s)	Gen.	Time/Gen.	B. dev.	CV(best)	A. dev.
SGAs	Nearest distance	40.591	56.8	0.71	89.00%	0.0646	169.29%
	Multi-criterion	325.406	76.8	4.24	28.52%	0.1395	59.85%
	Maximum flow	467.723	60.7	7.71	0.28%	0.0182	10.39%
	Minimum cost	498.632	62.8	7.94	0.06%	0.0124	8.13%
CGAs (75%)	Nearest distance	38.916	56.4	0.69	98.16%	0.1052	179.84%
	Multi-criterion	378.607	72.5	5.22	20.32%	0.0287	44.34%
	Maximum flow	398.081	78.2	5.09	0.18%	0.0029	9.26%
	Minimum cost	586.612	75.4	7.78	-0.08%	0.0030	7.98%

Table 5-17: Test results of different allocation patterns (H (300))

The best-know solution for Hub (300) is updated by instance CGAs (75%) with "minimum cost" allocation pattern to 62633239. For this reason, the corresponding value in Tab.5-17 is negative. The test results show that the demand allocation pattern of "minimum cost" performs the best, followed by the pattern of "maximum flow", and then by "multi-criterion" and "nearest-distance". Performance differences between them are proved to be significant by *T*-tests.

The test results conform to the conclusion by O'Kelly that the demand allocation pattern of "nearest-distance" ignores the flow between hubs so that it is not suitable to HLPs[429]. However, the performance of "multi-criterion" allocation pattern is not so good and even worse than that of "maximum flow" allocation pattern, which seems to be contrary to the conclusion of Klincewicz[430]. However, as we can see from Eq.4-5 and 4-6, these two greedy functions only consider the feeder transportation cost, while ignoring the expensive inter-hub air cost. Moreover, this kind of distortion may be aggravated in networks with unevenly-distributed demand nodes, since the multi-covered demand nodes can be aggregated in few hub regions, leading to higher air cost. In our case, and maybe also in most H/S systems, in which backbone travel cost is higher than feeder cost, the allocation pattern of "minimum cost" performs better.

Since the LS on demand allocation decision after CGAs (75%) is tiny, compared with running time for GAs, we neglect it when calculating the values of "running time per generation". The running time for algorithms with "nearest-distance" allocation pattern is significantly less than the other three. The calculation of exchange flow between demand node and potential "home" hub is quite time consuming. As we have anticipated, the running time for GAs with "minimum cost" allocation pattern is the longest.

When we take a deeper look at the solutions, we find that all solutions contain three hubs, although they may choose different hubs. It refers that the solution is not sensitive to hub number but hub location. We also find all solutions have a common feature that most of the demand nodes are still subordinated to one of the selected

[429] See O'Kelly (1987), pp.393-404.

[430] See Klincewicz (1991), pp.25-37. A multi-criteria assignment procedure to the p-Hub Median Problem were proposed by author: initial assignments for a given node i is based on a weighted sum of a common traffic measure with hub k (measured in units of traffic interchanged with nodes already assigned to k) and a distance measure to hub k (measured as the inverse distance from i to k , so that the closest hub has the largest measure).

hubs even with "minimum cost" allocation pattern. This may come from two reasons. One is that the hub radius is large enough to create large common area among different hub regions. As a result, a majority of demand nodes are multi-covered. The other reason is that demand nodes are unevenly distributed, which we have already mentioned in Sec.5.2.1. In order to save the expensive air cost, the model allocates as many as multi-covered nodes to one "home" hub. These two reasons may explain why most demand nodes are subordinated to one hub.

5.2.3.5 Overall improvement on SGAs

To the best of our knowledge, no other meta-heuristic has been proposed for our problem, whose result can serve as a benchmark for the hybrid GAs proposed in this dissertation. In this regard, we evaluate the performance of the improvement techniques under large-scale instances by comparing the results by the hybrid GAs with those by the untailored SGAs.

Part of the test results in Tab.5-17 can be used directly for the comparison, i.e. results by SGAs (with "multi-criterion" and "nearest distance" allocation pattern but without injection mechanism or LS after GAs) and results by CGAs (75%) (with injection rate 8%, "minimum cost" demand allocation pattern and FLS on demand allocation decision after CGAs). We choose the demand allocation patterns of "multi-criterion" and "nearest distance" for SGAs, since they are widely applied in HLPs and their performance has been verified by former studies[431].

	Allocation pattern	Running time(s)	Best solution	CV(best)	Average Solution
SGAs	Multi-criterion	325.406	80494934	0.0646	100118962
	Nearest distance	40.591	118376976	0.0646	168662861
Hybrid GAs	Minimum cost	586.612	62585891	0.0030	67632792

Table 5-18: Comparison between solutions by untailored SGAs and those by tailored hybrid GAs

As can be observed in Tab.5-18, the proposed hybrid GAs outperforms the two untailored SGAs significantly. The best solutions of hybrid GAs are 28.6% better than SGAs with "multi-criterion" allocation pattern and 89.1% better than SGAs with "nearest distance" allocation pattern, while the average solutions of hybrid GAs are 48% and 149.4% better than the two untailored SGAs respectively[432]. Moreover, the solutions generated by hybrid GAs are much more stable than those generated by untailored SGAs (see CV values).

The computational time required by hybrid GAs is about 9 times that of SGAs with "nearest distance" allocation pattern and 1.8 times that of SGAs with "multi-criterion" allocation pattern. But the longer computational time is justifiable since the proposed improvement techniques are able to improve the solution quality significantly.

[431] See e.g. Kratica (2007), pp.15-28; Klincewicz (1991), pp.25-37.

[432] The gap is calculated with this equation: gap = (SGAs- hybrid GAs)/SGAs.

The results of the comparison also conform to the "no free lunch theorems"[433]. Although meta-heuristics provide a generalized framework, its performance can be improved by embedding specific knowledge of the problem under consideration.

5.3. Summary of the test results

This chapter is dedicated to computational study on the overall performance of the proposed hybrid GAs under small-scale instances with CAB data set and the performance of proposed improvement techniques under large-scale instances with modified AP data set.

To evaluate the performance of the proposed hybrid GAs, we compare the results from hybrid GAs with optimal solutions generated by CPLEX. CPLEX (exact method) generally outperforms the hybrid GAs both in terms of running time and solution quality under instances with no more than 25 nodes.

To evaluate the performance of proposed improvement techniques under large-scale instances, we first modify the input data. Preliminary computational tests are then implemented to set efficient parameters for GAs. We reveal some relationship between Crossover probability, Mutation probability, running time, generation number and solution quality by comparing the test results under different instances. By comparing the running time of CPLEX and the hybrid GAs under small-scale instances and running time of the hybrid GAs under both small-scale and large-scale instances, we verify that the running time of exact algorithms grows exponentially, while meta-heuristics, such as GAs, experiences a moderate increase with the increase of instance scale.

The tests on the four improvement techniques not only reveal whether or how they can further improve the performance of the SGAs but also provide some general conclusions for GAs and other (meta-)heuristics for HLPs.

(1) Constructive procedure for initial solution generation

Constructive procedure borrow from GRASP is a quite effective measure to improve the solution quality of GAs. The randomness in constructive procedure brings diversification in the initial solution pool, while the greedy function leads to good solutions. Our test results prove that high quality and large diversity are desirable characteristics of initial solution pool for GAs. Although the running time for CGAs is longer than that for SGAs, the time is worthy for better solutions. However, a good parameter and a large solution space are prerequisites for the positive effectiveness of constructive procedure on GAs.

(2) Injection mechanism

Injection mechanism can improve the solution quality of CGAs by increasing the diversity of the solution pool during the process of GAs. CGAs with injection mechanism may converge faster than that without injection with certain constructive parameters. Meanwhile, the improvement effect of injection mechanism on SGAs is not obvious. Although the test results do not suggest an optimal injection rate for our test instance, they indicate that a moderate injection rate is usually preferred, while too high injection rate breaks up the balance between diversity and intensity, leading GAs to randomness.

---

[433] See Wolpert. Marready (1997), pp.67-82; Raidal (2006), p.3.

While constructive procedure increases the diversity of the initial solution pool, injection mechanism increases the diversity of the solution pool during the process of GAs. However, the test results indicate that the impact on the GAs performance from injection mechanism and constructive procedure is relatively independent and can hardly compensate to each other.

(3) Local search after GAs

The FLS on hub location decision after GAs does not change the best hub location decision. It may indicate that the time-consuming FLS does not work so efficiently after GAs with regard to the computational resource and the potential improvement.

Meanwhile, FLS on demand allocation decision after GAs can further improve the best solution with less time than FLS on hub location decision. We also find that the improvement by FLS on allocation decision is tiny if hub location decision is not improved. This consolidates the idea to apply PLS on allocation decision during the GAs, since it is not worth much effort to improve the demand allocation decision until the hub location decision is good enough.

(4) Demand allocation pattern before local search

The test results indicate that the demand allocation pattern of "minimum cost" performs the best, followed by the pattern of "maximum flow", and then by "multi-criterion" and "nearest-distance". The test results conform to the conclusion of O'Kelly that the allocation pattern of "nearest-distance" ignores the flow between hubs so that it is not suitable to HLPs. Moreover, the patterns of "multi-criterion" and "maximum flow" only consider travel cost in tributary network, while ignoring the more expensive air cost between hubs. They are also not suitable for H/S network with unevenly distributed demand nodes, otherwise the multi-covered demand nodes will be aggregated in few hub regions, leading to higher air cost. For air-ground H/S system, in which backbone cost rate is higher than feeder cost rate and demand nodes may be unevenly distributed, the "maximum flow" allocation pattern is more suitable, although the running time is a little bit longer.

Finally, we evaluate the overall performance of the four proposed improvement techniques by comparing the solutions by the hybrid GAs with those by untailored SGAs. Although the hybrid GAs requires longer computational time, it can provide better and more stable solutions. The results of the comparison also conform to the "no free lunch theorems" that the performance of meta-heuristics can be improved by embedding specific knowledge of the problem.

However, there are some flaws for our computational studies. The first one lies in the lack of benchmark to evaluate the performance of the proposed GAs under large-scale instances. Although we run the algorithm 10 times for each test instance and offer information about its stability in terms of CV values and reliability in terms of results of T-test, we do not know the deviation of the best-known solution by our algorithm from the optimal solution of the instance.

The second flaw is that we do not account for the interrelationship between parameters of GAs and those proposed improvement techniques, such as constructive process and injection mechanism, although we have made a lot of effort to find a relatively good parameter setting. In other words, we only check the performance of the proposed improvement techniques with a fixed GAs parameter setting. However, parameter setting itself is complicated enough to become a research topic[434].

[434] For study on parameter of GAs, reader may refer to Goldberg. See Goldberg (1989).

6. Empirical study on real-life problem

6.1. Input data preparation

The primary motivation of the proposed algorithm is to apply it to a real-life problem involving an EDS provider. As an application-oriented research, we try to provide our readers with an overview of the project, although not all the details are possible or permitted. We are aimed at not only theoretically solving the models by proposing some solution algorithms and improvement measures but also practically introducing some pertinent skills and techniques we have used during the implementation of the project. In Sec.1.3 we have already introduced how we position the target service with marketing instruments. The application of the proposed algorithm to our real-life problem requires comprehensive data collection, analysis and modification. In this section we introduce how we collect and modify the input data for the models. All data are compiled on daily basis.

6.1.1. Demand nodes set

After we position the target service- Next Morning and Next Day EDS [435], we should allocate limited resources and efforts to the right customers. In this section, we illustrate how the target market-demand nodes set *N*- is defined quantitatively, from which most high-end nationwide trans-city EDS demand can generate.

- **Methods and steps**

We consecutively employ a series of multivariate techniques, i.e. correlation analysis, principal component analysis (PCA) and cluster analysis (CA), to identify the target market. The pertinent methods and steps are illustrated in Fig.6-1.

[435] Please refer to Tab.1-2.

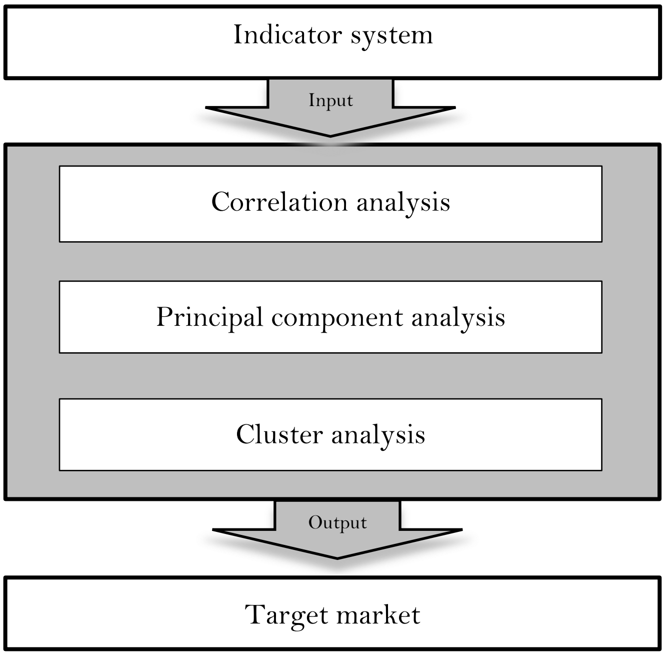

Figure 6-1: Methods and steps for identifying the target market

(1) Indicator system

Market analysis by the project committee is based on information from expert interviews, official publications and questionnaires handed out to corporate and private clients. Indicator system (see Tab.6-1) was designed accordingly in order to distinguish the high-end market under the consideration of the availability of the statistical data.

General indicators	Subordinate indicators
GDP	Contribution from secondary industry
	Contribution from tertiary industry
Export-oriented economy	Total export and import volume
	Foreign capital utilized
Freight volume	Highway freight volume
	Railway freight volume
	Air freight volume by civil aviation
Industrial economy	Industry output
	High-tech enterprises output
Commercial economy	Number of wholesalers and retailers
	Sales volume of consumer goods
Living standard	Disposable income of urban residents
Population size	Urban population size
Related business volume	Business volume of postal delivery service and EDS

Road traffic	Highway mileage

Table 6-1: Indicator system for identifying target market

(2) Correlation analysis

First of all correlation analysis is applied to check the co-linearity of variables with a correlation matrix table. The result of the correlation analysis by SAS shows that variables in the indicator system are highly correlated.

(3) PCA

PCA is therefore adopted to eliminate the co-linearity. It replaces the original difficult-to-interpret and correlated variables with fewer conceptually meaningful and independent components or factors in order to simplify the evaluation, while retaining most of the information in original data. In our case two independent indicators (or principle components in statistical terminology) abstract more than 80% of the information from the indicator system and explain most of the difference among cities (see App.3). The two component scores generated by PCA are taken as clustering variables.

(4) CA

CA is the kern step in the procedure illustrated in Fig.6-1. Many researches and books have pointed out that CA is a useful statistical instrument to identify group with similar characteristics[436].

CA, more specifically speaking, non-overlapping hierarchical method[437], we use here is a widely adopted method, which typically results in a dendrogram, i.e. a tree structure that represents the hierarchical relations among all objects being clustered. The dendrogram in Fig.6-2 is an example based on Ward's method (one kind of agglomerative clustering) with the modified project data set[438]. The *X*-coordinate represents the jointed cluster and the *Y*-coordinate indicates the loss of homogeneity[439]. The dendrogram is organized bottom-up that the merger of every possible cluster pair is considered and first minimizes the increase of within-cluster variance when groups are merged and then continues until all cities are clustered in one group.

Actually, clusters themselves are not directly derived by the hierarchical methods. Researchers seeking a solution with a certain number of clusters need to decide how to arrive at those clusters from the tree representa-

[436] See Churchill/ Iacobucci (2007), p 351; Myers /Mullet (2003), p.15; Wedel/Kamakura (1999) p.39. Books include such as McDonald/ Dunbar (2004) and Weinstein (2004).

[437] See e.g Everitt et al. (2011), Sec1.1; Romesburg (2004), p.2;

[438] Hierarchical structures can be basically derived from two types of algorithms: agglomerative and divisive methods. Agglomerative methods start with single-subject clusters, and proceed by successively merging those clusters at each stage of the algorithm until one single group is obtained. Divisive methods start with all subjects in one single group, and successively separate each group into smaller groups until single-subject groups are obtained. The latter category of methods is less popular in applied segmentation research.

[439] The semi-partial R-squared (SPR) measures the loss of homogeneity resulting from merging two clusters into a new one at each step. If the value is small, it suggests that the cluster solution derived at this step results from merging two very homogeneous clusters. On the other hand, large values of SPR suggest that two heterogeneous clusters have been merged to the new cluster.

tion. In our case a break point clearly occurs when three groups are merged into two, indicating the fusion of relatively dissimilar clusters after this point. If component scores of the two principle components from PCA for every city are projected in Euclidean space (see Fig.6-3), all the cities are also intuitionally divided into three groups. Other criteria, such as identifiability, accessibility and stablility[440] should also be considered for CA.

Figure 6-2: Process of cluster analysis

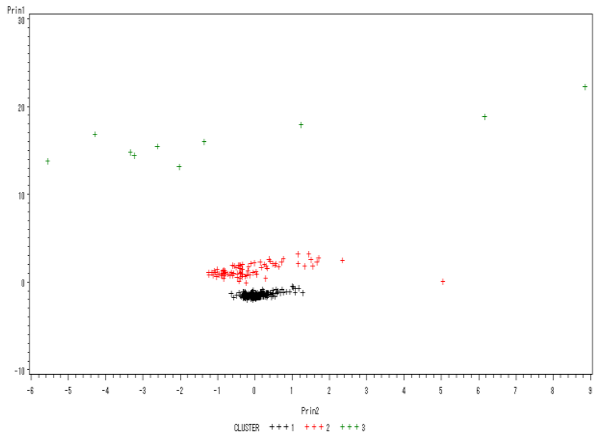

Figure 6-3: Component scores in Euclidean space

- **Results**

Every city in the database of Company A is regarded as a demand node. The methods and steps described above are adopted to divide all the more than 2000 cities into several groups. Analysis of variance (ANOVA) and hypothesis testing are executed on the result from CA. Finally, the first three groups that have the highest Component Scores are identified as target market (see Tab.6-2). Managers from Company A were quite satisfied with the results. All the crucial cites listed by experts from the headquarter of Company A are included in the target market. Moreover, the size of the target market (281 cities) is quite ideal for a market development strategy. Therefore, the 281 cities in the first three groups are defined as demand nodes set *N*. Descriptive Statistics for Different Customer Groups, please refer to App.4.

[440] Different criteria are mentioned by former studies. See e.g. Wedel/ Kamakura (1999), p4; Dibb (1999), p108; Tonks (2009), p343; Kotler/ Keller (2009), p 64. We explain some in the following.

Substantial: The segments are large and profitable enough to serve.

Accessible: The segments can be effectively reached and served. That is, they can be characterized by observably different means.

Differentiable: The segments can be distinguished conceptually and respond differently to different marketing-mix elements.

Stable: Only segments that are stable over time can provide the necessary grounds for a successful marketing strategy.

Familiar: To ensure management acceptance, the segments composition should be comprehensible.

Relevant: Segments should be relevant in respect of the company's competencies and objectives.

Compactness: Segments exhibit a high degree of within-segment homogeneity and between-segment heterogeneity.

Compatible: Segmentation results meet other managerial functions' requirements.

Group No.	Number of cities in group	Representative cities
1	10	Beijing, Chengdu, Guangzhou, Shanghai, etc.
2	82	Anshan, Baise, Baotou, etc.
3	189	Bengbu, Shaoxing, etc.

Table 6-2: Target market defined with statistical analysis

6.1.2. Potential hub set

In order to minimize the total cost of the network, we should include as many demand nodes as possible in the potential hub set *H*. Moreover, studies on EDS networks in the USA show that it is common for EDS providers to locate their hubs in relatively small cities, where there are few other airlines competing for nighttime runway access[441]. For these reasons, what we did beforehand was to delete the cities without qualified airports from the demand node set *N* and include the rest cities in the potential hub set *H*. By this way, there are 187 nodes in the potential hub set *H*, including 5 nodes that are predetermined as hubs in the new network by managers, since Company A has already made intensive investment on these 5 hubs and has no intention to close them. In this respect, there are 182 nodes to be considered during the solution.

However, managers from Company A, who were in charge of the business, were reluctant to accept our method. They brought forward two reasons for their rejection. First and foremost, the location of hubs is not only a matter of total cost but also a matter of the market share of next-morning EDS, which is the premium service the to-be-planned network supports. As only hub cities can be offered with this premium service, locating hubs in small and less developed cities will result in losing the corresponding market share tremendously. By constraining hubs in relatively large and more developed cities, the total cost of network may be a little bit higher, but the revenue can be much higher with higher price for the premium service. As a matter of fact, this is one of the deficiencies in our models. We only concern the cost of the network, while neglecting the corresponding service revenue and other marketing considerations. The second reason is the trouble and difficulty to collect so much information about hub fixed cost for different cities. Meanwhile, the managers believed the result from CA provided information for the choice of candidate hubs, i.e. potential hub set *H* can be created on the basis of cities in Group1 and Group 2 in target market, which have larger index values than cities in Group 3 (see Tab.6-2).

Under the consideration of their persuasive concerns, we build another potential hub set *H*. We name it *Hs* and the potential hub set with our method *Hl*. For *Hs*, potential hubs are firstly confined to the 92 cities in Group1 and Group 2. 17 cities are then eliminated due to the deficiency of qualified airports. 75 cities in Group 1 and Group 2 are included in potential hub set *Hs*.

However, the small size of the data set *Hs* leads to a new problem that not all the demand nodes in *N* can be covered by potential hubs in *Hs*. We would like to introduce how we treat this problem. First, for every demand node in *N* that itself is not a potential hub, we scan if there are potential hubs within its coverage. If there is more than one, we do nothing. If there is only one, we designate it as hub during the solution process. If there is no potential hub within its coverage, we take 5 top demand nodes in CA output (if exist) to the man-

[441] See Kuby/ Gray (1993), p.10.

agers and keep at most 3 of them in the potential hub set *Hs*. If there is neither potential hub nor other demand node within its coverage, we designate it as hub. After this modification, *Hs* contains 125 nodes.

Finally, we present the managers both solutions with potential hub sets *Hs* and *Hl* . The difference between the counterparts explains how the managerial constraint impacts on the solutions of the problems. The corresponding cost discrepancy can be used to evaluate the cost of the market strategy and the price of the premium service.

We use *Hl* with 187 potential hub nodes in the following for test purpose as we believe large size of potential hub set can help us to reveal the relationship between different influence factors and solutions more thoroughly.

6.1.3. Demand volume forecast

Another essential input of the models is demand volume between every O-D pair in target market. Demand forecast itself is an important research topic in this project. We briefly introduce the methods we use.

Since the key service defined in Tab.1-2 is quite new for Company A, there is no historical business data for reference to forecast the demand volume for the network planning. Even historical business data of similar service for each O-D pair is not available. In this respect, demand forecast of this brand new service system has to be based on aggregate demand forecast for EDS from individual city. Specifically, factors that can have impact on EDS demand are identified. Time series analysis is applied to derive aggregate demand for EDS. The percentage of high-end demand and also the market share by Company A on this high-end market are also estimated to derive the demand forecast for particular service (see Fig.6-4).

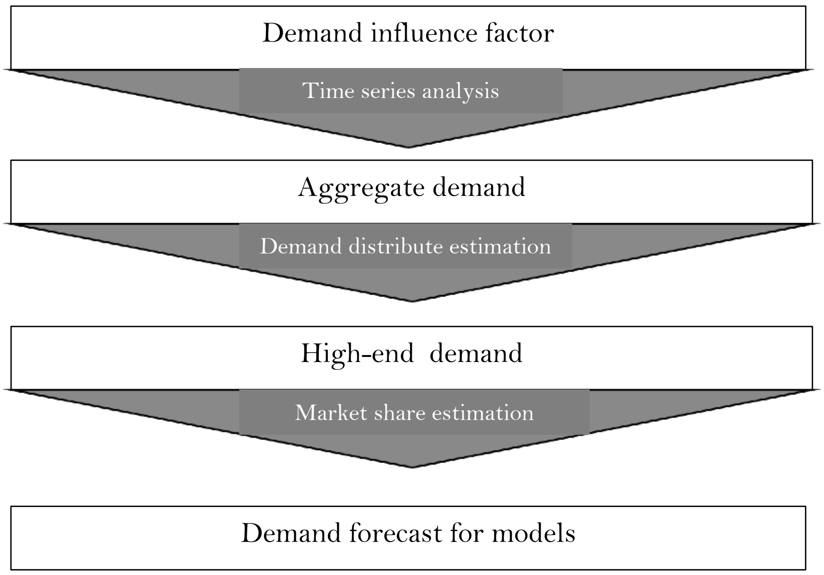

Figure 6-4: Methods and steps for demand forecast

6.1.4. Daily time schedule

According to Fig.3-3 and Fig.3-4 in Section 3.1.2, there are 9 or 7 activities along the parcel route from origin to destination. We list a typical time schedule for path "normal demand node to normal demand node" and a corresponding one for path "in-hub demand node to in-hub demand node" in Tab.6-3. It is used in tests in Sec6.2 as basic instance and also serves as the benchmark for the scenario planning in Sec.6.3.

No.	Activity	Normal to normal		In-hub to in-hub	
		Time period	Time duration	Time period	Time duration
1	Pickup from shipper	Until 18:30	Whole business day	Until 21:30	Whole business day
2	Twilight local sortation	18:30-19:30	1h	21:30-22:30	1h
3	Feeder transportation	19:30-22:30	3h	-	-
4	Twilight regional sortation	22:30-24:00	1.5h	22:30-24:00	1.5h
5	Air transportation to destination hub	0:00-5:00	5h	0:00-5:00	5h
6	Sunrise regional sortation	5:00-6:30	1.5h	5:00-6:30	1.5h
7	Feeder transportation to destination city	6:30-9:30	3h	-	-
8	Sunrise local sortation	9:30-10:30	1h	6:30-7:30	1h
9	Distribution to consignee	From 10:30		7:30-10:30	3h

Table 6-3: Time window of path from "normal demand node" to "normal demand node" and from "in-hub demand node" to "in-hub demand node"

Highway system is unevenly developed in different areas in China. We can define different coverage radii for potential hubs. An alternative way is to define different cut-off time and arrival time, i.e. different service policies in different regions. For example, the local cut-off time in less developed regions with unsatisfactory road condition is earlier than that in developed regions with advanced highway system. In our test instances, we simply assume that all hubs have the same coverage radius in terms of distance. We also assume that the average speed for feeder trucks on highway is the same in all hub regions- 90km/hour. Consequently, the time window for feeder transportation in all hub regions and the local cut-off time for all normal demand nodes are the same.

6.1.5. Air cost by self-owned aircraft

Company A has its own air cargo daughter company that owns and financing leases several different types of aircraft. Air cost by self-owned aircraft is estimated with sum of fixed cost and variable cost.

Fixed cost is attributed to the ownership and maintenance of the aircraft. It includes all costs that are independent of traffic volume, such as the purchase cost, crew cost, taking-off and landing fees, parking cost and repair cost. Fixed cost would differ from aircraft to aircraft and depend on factors such as the age of the aircraft, kilometers flown, etc. The fixed cost in our models is estimated in cost/kilometer (see k_s in Tab.3-5). First, the daily fixed cost of an aircraft is calculated with the purchasing cost for the brand new aircraft with a 20-year span depreciation and the other costs mentioned above. Then the average flown distance per day of that aircraft is estimated based on historical data. Finally, the fixed cost per kilometer k_s is calculated by dividing daily fixed cost with daily average flown distance. In this way, the fixed cost attributed to each inter-hub link is proportional to the length of the link.

Variable cost χ_s is attributed to the operational cost of aircraft that depends on the traffic volume, take fuel as an example. It is estimated in cost/ (kilo* kilometer). We approximate the average variable cost of an aircraft of type s with company's one-year historical data by Eq.6-1. The numerator summates the variable cost of an aircraft of type s in one year. Parameter v_s represents the traffic volume that aircraft actually carries and d_s the distance it flies. The numerator and denominator are calculated for the same time period. For the sake of simplicity, in the following computational tests we assume that both the fixed cost and the variable cost for one type of aircraft are the same.

$$\chi_s = \frac{\mathrm{var}_s}{v_s \times d_s} \tag{6-1}$$

6.1.6. Capacity constraint on self-owned aircraft

Every type of aircraft has a technical maximum capacity. We modify the capacity constraints in the models with an average payload rather than use the technical maximum capacity directly in order to consider demand randomness and seasonal fluctuation.

We enforce a payload reserve margin for accommodating above average daily demand (see Eq.6-2). More accurate method is to derive a deterministic volume by taking, e.g., 90th percentile as the volume that must be planned for. However, there are no historical data to simulate a convincing demand distribution. We just estimate the over-demand risk roughly with a fixed probability, i.e. payload. A higher payload means lower cost but higher risk.

$$u_s^u = technique\ \ capacity_s \times average\ \ payload \tag{6-2}$$

For company A, the value of this payload is not a critical concern in the network planning. There are other ways and mechanisms Company A can use to evade the over-demand risk and save cost simultaneously. On the one hand, these self-owned aircraft are in practice shared by other services by Company A, such as economical express and logistic service. So the overall transportation cost at corporate level is not so sensitive to the payload. On the other hand, during the busy season, such as Chinese Rural New Year, it is also practicable to lease shipping space from air freight market. For these reasons, both the risk and cost is not so sensitive to the payload in our case.

The average payload is set here as 80% for test purpose[442]. Since our special case is not applicable to normal situation, we will also test how the average payload can impact on the network structure and corresponding total cost.

[442] In passenger airline, this parameter is set to be lower than that in air freight application. For example, Akyin set it as 60% for his computational test. See Aykin (1994), p. 516.

6.1.7. Illustration of other input data

- Flight time T_{km} and feeder transportation time T_{ik}

The flight time between any potential hubs T_{km} $(k, m \in H)$ is provided by the Civil Aviation Administration of China. One can also roughly estimate it by dividing the flight distance with an average speed of the aircraft[443]. With the help of an intelligent navigation system in vehicles, we easily get the latest data for the highway distance between any two demand nodes d_{ik} $(i \in N, k \in H)$. Feeder transportation time T_{ik} is then calculated based on an average speed of trucks on highway according to experts' experience. We set it as 90km/h here for the test purpose.

- Cost rate on air freight market

Air cost from the freight market is estimated in cost/ (kilo* kilometer). It is subject to many factors, e.g. competition on regional markets, fuel price, seasonal demand, etc. It suffers a high fluctuation especially in developing countries like China. We estimate the air cost from the freight market based on the average cost rate by self-owned aircraft in Company A and an average profit margin in air freight industry.

- Feeder transportation cost rate γ

Feeder transportation cost rate is also estimated in cost/ (kilo* kilometer). In China a substantial part of the overland freight service is offered by third-party truck fleets. Market rate of overland transportation suffers a high fluctuation due to the fast variation of demand and supply. However, Company A has long-term agreements with some regional truck companies, who charge a stable contract rate all year round.

- Hub fixed cost fh_k

Hub fixed cost is incurred, when a new hub is to be established or an existing consolidation center is to be expanded. We disregard the residual value of all the existing consolidation centers with Sunk Cost Theory. When an existing consolidation center is selected as hub in the new network, only the corresponding expanding cost is calculated in the model. In this respect, the new network gives some preference to the existing facilities.

The fixed cost for establishing or expanding a current hub at a given city includes the cost for labor, equipment, maintenance, and, most importantly, the cost of land, which largely depends on the policy of the local government. Daily hub fixed cost is estimated accordingly based on some assumptions on amortization parameters, such as discount rate and lifetime of the network.

Nevertheless, in the models we neglect the hub variable cost, i.e. the parcel handling cost. The total parcel handling cost is not so sensitive to the decisions on network structure (hub location and demand allocation

[443] It is an approximation for reality, since cruising speeds for different aircraft are not the same. Mahapatra S. described in detail how to estimate flight time by using a regression model. See Mahapatra, (2005), pp. 34-36.

decisions) according to our understanding of the problem. It is not worth the computational effort to consider the tiny cost discrepancy in the models.

6.2. Computational tests under basic instance and results analysis

In this section, we carry out computational tests on *Ext.1* and *Ext.2*, provide the corresponding results and make some analysis and comparisons.

6.2.1. Result of *Ext.1*

- **Specifications of algorithm parameters and input data**

The solution of extension models follows the process illustrated in Fig.4-3 in Sec. 4.1.3. Air service selection decisions in *Ext.2* are derived with the Integer Programming Toolbox embedded in Matlab.

We make similar computational tests as those in Sec.5.2.2 to set the algorithm parameters for the project data set. We also selectively apply the improvement techniques in Sec.4.2 based on the test results. The algorithm for the project data set in this chapter is run with CGAs (50%), Population size N_{pop} 50, Maximum generation $N_{\max\_gen}$ 200, Minimum generation $N_{\min\_gen}$ 50, Injection rate P_{new} (8%), Maximum number of iteration without 0.001% improvement N_{imp} 5, Crossover probability P_{cro} 60%, and Mutation probability P_{mut} 20%.The starting point of PLS on allocation decision embedded in CGAs is determined by "minimum cost" allocation pattern" and the hybrid CGAs is followed by a FLS on allocation decision.

We make computational tests on extension models in this chapter with modified demand and cost data based on the project data set out of the confidential reason, while keeping reasonable interrelationship among the input data. In Tab.6-4 and Tab.6-5 we list input data for the basic instance, which also serves as benchmark for scenario planning in Sec.6.3.

Input data	Value
Demand nodes in N	281
Potential hub nodes in H	187
Hub fixed cost fh_k (daily)	500000
Cost rate for feeder transportation by truck γ	0.005/(kilo*km)
Hub coverage radius D	270km

Table 6-4: Input data of the basic instance for extension models

We consider five types of air service in *Ext.1* (see Tab.6-5). Service Type 1 denotes service from air freight market, while Service Type 2, 3, 4 and 5 denote service by self-owned aircraft. Since Service Type 2 is not included in current aircraft fleet of the company, only Service Type 1, 3, 4, and 5 are considered in *Ext.2*.

Nr.	Fixed cost k_p (/km)	Variable cost χ_p (/kilo*km)	Lower bound u_p^l (kilo)	Upper bound u_p^u (kilo)
1	0	0.0070	0	∞
2	0.3	0.0068	0	2700
3	1.5	0.0065	0	5000
4	3.5	0.0060	0	8000
5	14	0.0055	0	15000

Table 6-5: Cost functions for different air services

- **Stability of the results**

We still apply the index of CV (see Eq.5-1) to indicate the stability of the results under with the project data set. We list the best solutions of 10 runs with *Ext.1* under the basic instance in Tab.6-6 and provide the corresponding CV value at the bottom. The deviations from the best-known solution are calculated with Eq.5-4.

Test number	Best solution		Running time (s)
	Value	B. dev. (%)	
1	56699837.84	0.543%	1727.81
2	56515604.09	0.216%	1689.54
3	56393762.12	0.000%	1758.65
4	56393762.12	0.000%	1711.98
5	56699837.84	0.543%	1738.42
6	56515604.09	0.216%	1750.21
7	56393762.12	0.000%	1758.94
8	56393762.12	0.000%	1699.45
9	56515604.09	0.216%	1702.26
10	56393762.12	0.000%	1786.38
Average	56491529.86	0.17%	1732.36
Standard deviation	123069.89		
CV	0.0022		

Table 6-6: Summary of the solutions from 10 runs with *Ext.1* under the basic instance

- **Best-known solution with *Ext.1* under the basic instance**

Result	Value
Total cost	56393762
Air cost	19409702
Feeder transportation cost	2484060
Hub number	69
Volume by self-owned aircraft (kilo*kilometer)	1146411475
Volume by air freight market (kilo*kilometer)	1657526136

Volume by truck (kilo*kilometer)	496812000
Hub links by service type 1	4455
Hub links by service type 2	193
Hub links by service type 3	0
Hub links by service type 4	44
Hub links by service type 5	0

Table 6-7: Best-known solution with *Ext.1* under the basic instance

6.2.2. Computational tests on Improvement technique 5 and result of *Ext.2*

- **Specifications of the numerical constraints on air service**

We take up the same solution process, algorithm parameters and input data as those for *Ext.1* to solve *Ext.2* but with additional numerical constraints on each service type according to the company's current aircraft fleet (see the last column in Tab.6-8).

Nr.	Fixed cost k_p (/km)	Variable cost χ_p (/kilo*km)	Lower bound u_p^l (kilo)	Upper bound u_p^u (kilo)	Numerical constraints
1	0	0.0070	0	∞	∞
2	0.3	0.0068	0	2700	0
3	1.5	0.0065	0	5000	2
4	3.5	0.0060	0	8000	10
5	14	0.0055	0	15000	15

Table 6-8: Numerical constraints on air services with *Ext.2*

- **Computational tests on Improvement technique 5 for *Ext.2***

To test if Improvement technique 5 proposed in Sec.4.2.5 can work positively under our real-life instance and if it is as efficient as we expect, we modify the program to get the information we need. That is, after the reproduction procedure, i.e. crossover and mutation operation, in each generation, we also calculate the average solution of the solution pool with an estimated average air cost rate - 0.0069/ (kilo*km). The effect of the proposed technique is analyzed based on the average solution of the solution pool, since we pay more attention to the convergence of the solution pool that determines the termination of the algorithm, while the best solution may already be generated in earlier generation.

We make one probe test with the integer programming for the air service selection decision invoked for every individual in each generation from the beginning of the hybrid GAs, i.e. initial solution generation procedure. With the modified program we get two data sets, one for average solutions with exact air cost and the other for average solutions with approximated air cost. We denote $Avg._f^E$ the average solution with exact air cost in the former generation, $Avg._c^E$ that in current generation and $Avg._c^A$ the average solution with approximated air cost in current generation. The improvement on the average solution by the overall algorithm and that by the integer programming of each generation is calculated with Eq.6-3 and Eq.6-4 and spotted in Fig.6-5. Note that the latter one is actually the rectification by the integer programming to the approximation.

$$\text{Imp.overall}=\frac{Avg._f^E - Avg._c^E}{Avg._f^E}\times 100\% \tag{6-3}$$

$$\text{Imp.IP} = \frac{\left|Avg._c^A - Avg._c^E\right|}{Avg._f^E}\times 100\% \tag{6-4}$$

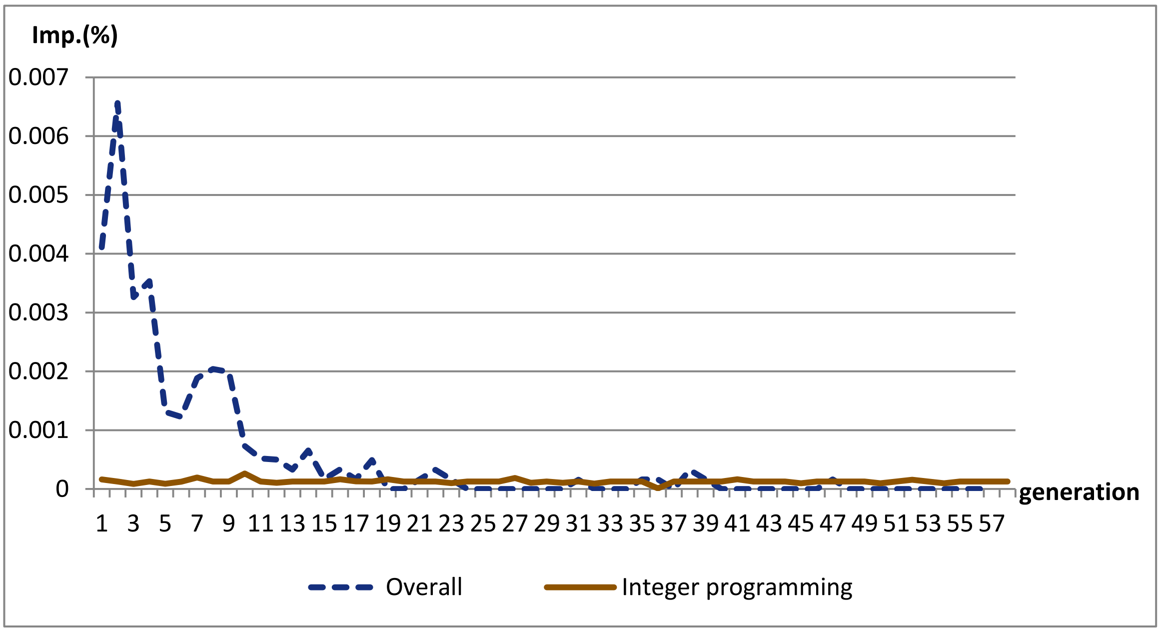

Figure 6-5: Improvement on the average objective value of the solution pool by the overall algorithm and that by the integer programming

When we compare the improvement on average solution by the overall algorithm and that by the integer programming, we find that the former one is much larger than the latter one in early stage and comes closer to the latter one gradually, while the latter is quite stable through the whole process. As we have mentioned before, the optimal air service selection decisions largely depend on hub location and demand allocation decisions. According to our understanding of the problem, it makes no sense to find the optimal solutions of air service selection problem on the basis of bad location and allocation decisions. In our case, the location and allocation decisions play a decisive role on the solution quality.

With regards to this, we can insert the integer programming for air service selection decisions after the solution pool of the hybrid GAs is relatively stable, e.g. after the 10th generation in our case. Before that, the air cost can be estimated with an approximated average cost rate, e.g. a cost rate that is a little bit lower than the market cost rate. In other words, the algorithm before certain generation runs with the basic model and it changes to *Ext.2* after that.

In order to test the performance of the proposed technique, we run the algorithm with the improvement technique (called improved algorithm or *IAlg.* for short) and without (called normal algorithm or *NAlg.* for short)

10 times for each. In *IAlg.* the air cost is approximated with an average rate as 0.0069/ (kilo*km) and the integer programming is invoked by GAs after the 10th generation. The best-known solution among these 20 runs is 56540386. The running time, the deviation of the best solution from the best-known solution (denoted as B. dev. and calculated with Eq.5-4) and the deviation of the average solution from the best-known solution (denoted as A. dev. and calculated with Eq.5-4) are recorded in Tab.6-9. The CV values at the bottom are based on original solutions and expressed not in terms of percentage.

Generation No.	**Running time(s)**		**B. dev. (%)**		**A. dev. (%)**	
	IAlg.	***NAlg.***	***IAlg.***	***NAlg.***	***IAlg.***	***NAlg.***
1	1703.457	1705.864	0.0000	0.7480	0.8527	1.5924
2	1758.984	1723.119	0.7480	0.0019	1.5939	0.8518
3	1678.574	1727.809	0.0000	0.0019	0.8528	0.8530
4	1723.098	1793.847	0.0019	0.0000	0.8529	0.8537
5	1755.483	1744.563	0.7480	0.0024	1.5932	0.8521
6	1728.465	1699.485	0.0000	0.0024	0.8533	0.8534
7	1764.768	1739.089	0.0019	0.0000	0.8529	0.8551
8	1695.365	1782.967	0.7480	0.0000	1.5928	0.8565
9	1691.376	1809.023	0.0000	0.0024	0.8531	0.8531
10	1745.987	1768.675	0.0000	0.0000	0.8533	0.8527
Average	1724.5557	1749.444	0.2248	0.0759	1.0751	0.9274
CV (not in %)	0.0181	0.0215	0.0036	0.0024	0.0035	0.0023

Table 6-9: Performance comparison between *IAlg.*and *NAlg.*

IAlg. finds the best-known solution 5 times in 10 runs, while *NAlg.* finds 4 in 10 runs. Both the best solution and the average solution in the final solution pool of *NAlg.* are better than those of *IAlg.* on average (see the average values in last row of Tab.6-9). However, the results of the T-tests on both best solutions and average solutions of the two algorithms are not significant, which means the solution quality of *IAlg.* is basically the same as that of *NAlg.* Moreover, the solution stability of the two algorithms is similar in terms of the CV values for both best solutions and average solutions. In other words, the negative impact on the solution quality by the approximation technique is not significant.

However, the result of the T-test on running time also shows that there is no significant difference between the two algorithms, although the average running time of *IAlg.* is about 25s shorter than that of *NAlg.* In this case, we cannot provide convincing evidence to prove that the proposed technique can save solution time, although it also has no evident negative impact on the solution quality. It may result from two reasons, neither of which we can provide evidence to prove. One reason may be that our approximation is not so good that the GAs have to run more generations to adjust this bias after we insert integer programming. The other more convincing reason is that the randomness of the running time from the hybrid GAs covers up the time-saving effect of the proposed technique. The problem scale of the integer programming in our case is relatively small and its running time by the Integer Programming Toolbox embedded in Matlab is relatively short. Meanwhile, the average running time per generation of the hybrid GAs is relatively long, i.e. about 30s. Even small randomness of the generation number can balance out the time-saving effect of the proposed technique.

Nevertheless, theoretically the proposed technique can achieve its deserved effect, but when (1) a good approximation can be made and (2) the embedded exact algorithm is time costly compared to the master heuristics.

Compared with *NAlg*, *IAlg*. has not obvious advantage in our case. So we continue to use *NAlg*. in the following computational tests.

- **Result of *Ext.2***

The CV value for the 10 runs with Ext.2 is 0.0024 (see the last line of Tab.6-9). As we can anticipate, the stability of the algorithms for *Ext.1* and *Ext.2* should be essentially the same, since the embedded integer programming is solved optimally when invoked. In Tab.6-10 we present the best-known solution with *Ext.2* under the basic instance.

Result	**Value**
Total cost	56540386
Air cost	19555555
Feeder transportation cost	2484831
Hub number	69
Volume by self-owned aircraft (kilo*kilometer)	236898338
Volume by air freight market (kilo*kilometer)	2567039273
Volume by truck (kilo*kilometer)	496966110

Table 6-10: Best-known solution with *Ext.2* under the basic instance

6.2.3. Comparison of results of *Ext.1* and *Ext.2*

In this section we compare the results of *Ext.1* and *Ext.2* to reveal some features of the planned network. Also we check whether and how the numerical constraints on self-owned aircraft can impact on the network structure, total cost, inter-hub flow, etc.

- **Hub location**

The best-known solutions of both extension models choose the same 69 nodes as hubs, indicating that the hub location decision is not sensitive to the imposed numerical constraints on self-owned aircraft.

Moreover, they are also the solutions with the least hub number in the final solution pool. This may indicate that the high hub fixed cost drives the model to choose as few hubs as possible to balance the high fixed cost with the transportation cost (see Fig.6-6). It is supported by the overall cost distribution illustrated in Fig.6-7 and Fig.6-8, in which the hub fixed cost is much higher than the transportation cost. p' in Fig.6-6 represents the optimal hub number in our models, while p^* represents the optimal hub number in the corresponding hub location problem with fixed cost[444]. Subject to the hub coverage radius constraints, our models have to locate more hubs than the corresponding model that has no coverage constraints.

[444] It has no constraints on hub coverage radius. For details, please refer to Sec.2.2.4.

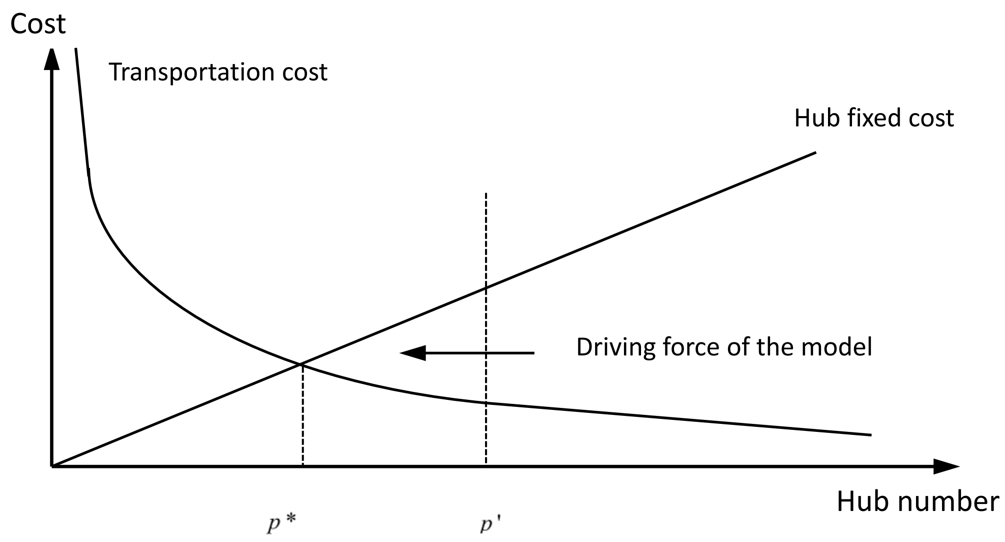

Figure 6-6: Optimal hub number

- **Cost distribution**

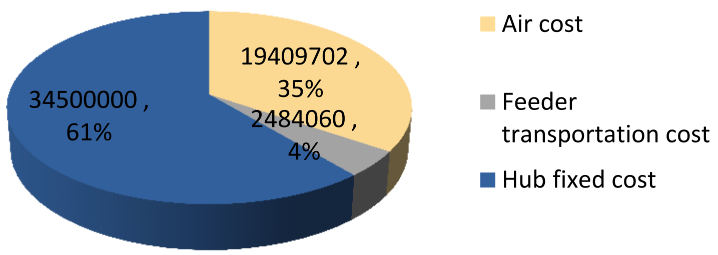

Figure 6-7: Cost distribution over the network under the basic instance (*Ext.1*)

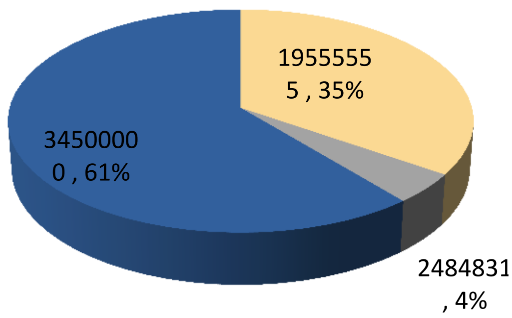

Figure 6-8: Cost distribution over the network under the basic instance (*Ext.2*)

From Tab.6-7 and 6-10 we can easily calculate the cost distribution over the network in *Ext.1* and *Ext.2* (see Fig.6-7 and Fig.6-8). In both models the hub fixed cost takes up more than 60% of the total cost. Contrarily to pure ground H/S networks, the air-ground H/S network here has much higher transportation cost for the backbone network than that for the tributary networks. The high air cost results from the high cost rate of air service and long flight distance compared to the small hub coverage radius. These are the reasons why we pay more attention to the air network, while simply assuming star-shaped tributary networks with an average cost rate.

Without numerical constraints on self-owned aircraft, *Ext.1* certainly incurs lower air cost than *Ext.2* by optimally selecting air service on each hub link. Moreover, the feeder transportation cost in *Ext.1* is also lower than that in *Ext.2*. Maybe some demand nodes in *Ext.2* turn to hubs that are connected with more self-owned aircraft to reduce air cost.

- **Traffic volume**

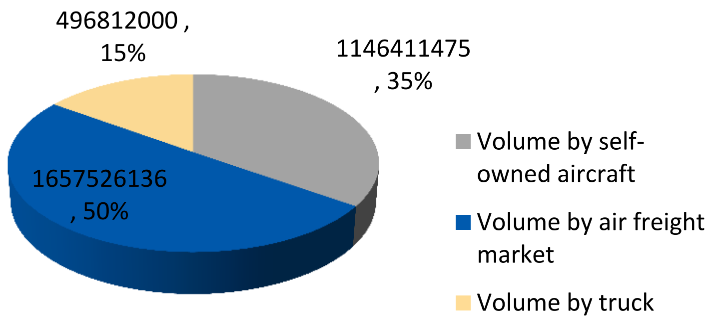

Figure 6-9: Traffic volume under the basic instance (*Ext.1*)

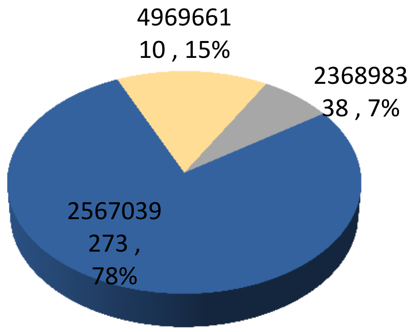

Figure 6-10: Traffic volume under the basic instance (*Ext.2*)

We go over the traffic volume by different transportation modes and air service in both *Ext.1* and *Ext.2* (see Fig.6-9 and 6-10). Although the majority of the demand must be fulfilled with road transportation (except the demand from in-hub node to in-hub node), traffic volume by truck (in kilo*kilometer) takes up only 15% of the total volume in both models. It is believed that the small hub coverage radius due to the tight time constraints shifts the cost focus of the network from tributary networks in pure ground H/S networks to the backbone air network in air-ground H/S networks. It is an inevitable trend of the network for EDS, also indicating that the planning focus of multimodal EDS networks should also lie in the air network.

The result of *Ext.1* (see Fig.6-7) shows that 41.18% (35%/85%) of the total air freight is fulfilled by 5.05% [44+193/4692] of the flights operated by self-owned aircraft. Similar phenomenon can also be distinguished in result of *Ext.2* (see Fig.6-10) with 8.24% (7%/85%) of the total air freight by 0.575% [(2+10+15)/4692)] of the flights. In this respect, it can be inferred that the distribution of the inter-hub flow is strongly uneven.

- **Inter-hub flow**

Uneven distribution of the inter-hub flow occurred frequently in previous studies on HLPs[445], especially in networks with concave cost functions[446]. That is, relatively large traffic goes through few highly discounted inter-hub links, while other normal inter-hub links have low traffic volume.

[445] See e.g. O'Kelly (1987), pp.393-404; Campbell (1994a), pp.387-405; Skorin-Kapov et al (1996), pp.582-593.

[446] See e.g. O'Kelly/Bryan (1998), pp.605-616; Klincewicz (2002), pp.107-122.

In order to investigate the incentive of flow bundling by the cost selection function, we compare the inter-hub flows in *Ext.1*, *Ext.2* and the basic model. The average air cost rate in the basic model is set to be 0.00692/(kilo*kilometer), which is the average air cost rate with *Ext.1* by dividing the total air cost with the total air traffic volume. Since the hub location decision is the same in these three models, the flow bundling effect by the cost selection function can be investigated by comparing inter-hub flows in these three models.

We go over the matrices of inter-hub flow with the three models and present the frequency lying within different intervals in Fig.6-11. It demonstrates that the inter-hub flows with the three models are all unevenly distributed with a unilateral long tail. In particular, more than 90% of the inter-hub links have less than 1000 kilos traffic volume, while the largest traffic volume reaches 12722 kilos.

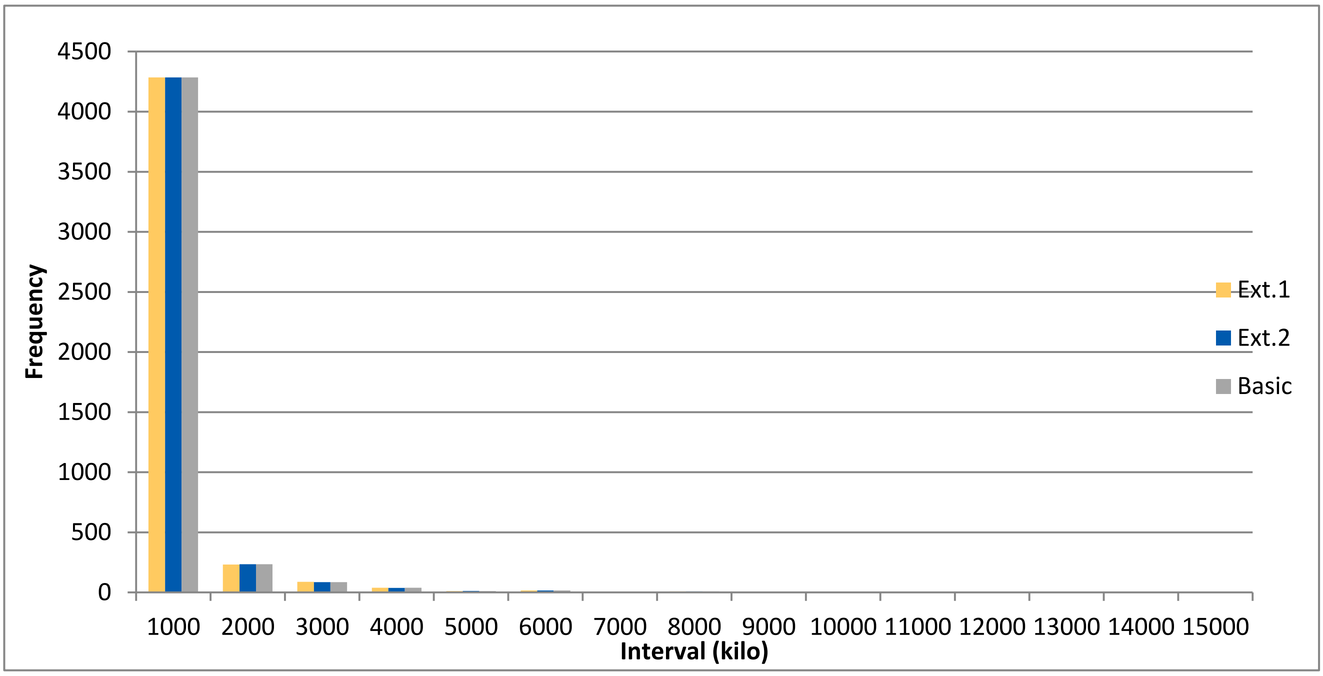

Figure 6-11: Frequency statistics of inter-hub flows with *Ext.1, Ext.2* and the basic model

The nature of the project data set is one of the reasons for the flow bundling, since the result of the basic model also shows the flow bundling. In the following we will also show that the distribution of the demand nodes in the network is also uneven.

Meanwhile, the cost selection function also has flow bundling effect. As we can see in Fig.6-11, there are a few discrepancies between the basic model and extension models within the interval of [1000-5000]. Since the hub location decision is the same, the differences on inter-hub flow between the three models totally result from the cost selection function and the numerical constraints on air service.

In extension models there are some opportunities to pass through the critical point to the next piece of cost function with lower rate by amassing flow on a single inter-hub link. For example, because of the Service Type 2 *Ext.1* has 89 inter-hub links with traffic volume between [2000, 3000], 3 more than the basic model and *Ext.2*. We can anticipate that the extension models, which are embedded with a cost select function, force some interacting pairs to utilize inter-hub links with lower cost rate to minimize the total transportation cost even though those are not the least-cost paths for these interacting pairs.

However, the bundling effect in our extension models is not so strong as that in former studies with concave cost function, such as FLOWLOC[447]. It may come from three reasons.

(1) The granularity of frequency in Fig.6-11 is not small enough to present all detailed differences between the three models. Indeed, the traffic volume on some inter-hub links in extension models is larger than the volume on the corresponding links in the basic model, but not large enough to go across the breaking points in Fig.6-11.
(2) Constraints on hub coverage radius decrease the opportunity to reallocate demand nodes for agglomerating the flow. As a matter of fact, 137 out of the 281 demand nodes are covered by only one hub under the basic instance with hub coverage radius 270 km. In other words, there are only 144 demand nodes can be reallocated.
(3) Single allocation policy weakens the flow bundling effect. It is found that under the multiple allocation policy an increase in the number of single allocation occurs as the inter-hub discount increases[448]. So the bundling effect in single allocation network is not as evident as that in multi-allocation network with the same cost incentive.

- **Hubs and subordinate demand nodes**

It is demonstrated by the demand allocation decisions with the two extension models that demand nodes are agglomerated in few hub regions rather than evenly allocated to all hub regions. Fig.6-12 displays how the 281 demand nodes are allocated to the 69 hub regions. The *X*-coordinate denotes the number of demand nodes in a hub region, while the *Y*-coordinate denotes the frequency.

[447] See O'Kelly/Bryan (1998), pp.605-616.

[448] See O'Kelly (1998), pp.171-186.

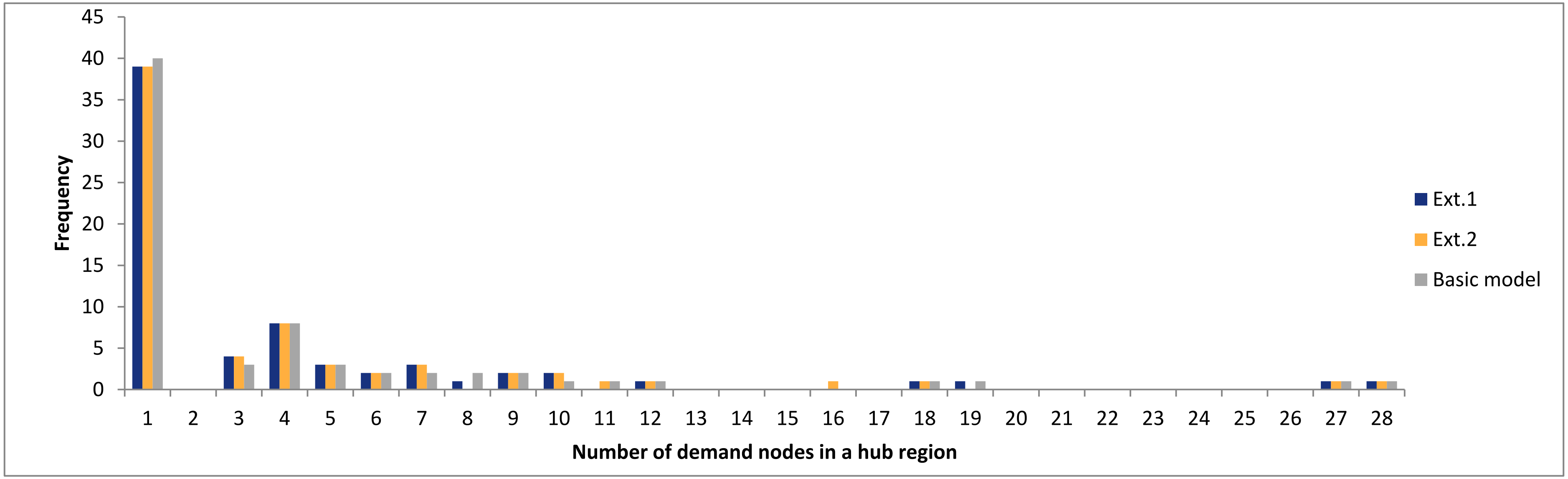

Figure 6-12: Demand nodes distribution among hub regions

Note that demand nodes are unevenly allocated in all the three models, i.e. 39 of the 69 hubs have no subordinate normal demand node but only one in-hub demand node, while the largest hub region has 28 subordinate demand nodes.

The agglomeration of demand nodes results from two reasons, i.e. the cost select function and the regional economics in China. We also believe that the latter one plays a decisive role, since the demand nodes in the basic model are also seriously unevenly distributed. Promoted by national policies, strong economic relationship has been developed in some regions, i.e. economic circles. The results of all the three models obviously confirm and verify the phenomenon of regional economics in China. The five largest hub regions in all the three models (see Fig.6-12) correspond exactly to the five most economically developed regions in China[449].

Hub region	**Total demand nodes**	**Covered demand volume (%)**	**Involved self-owned aircraft**
9	18	11.07%	2
23	28	12.59%	15
41	27	7.84%	10
49	19	6.68%	7
169	12	5.37%	2
Sum	104	43.55%	36

Table 6-11: Brief summary of the five major hub regions (based on *Ext.1*)

We make a brief summary of the 5 major hub regions based on the solution with *Ext.1* (see Fig.6-11). The index “Total demand nodes” denotes the number of demand nodes in the corresponding hub region. The 5 major hub regions cover 37.37% of the total 281 demand nodes. “Covered demand volume (%)” is calculated by summing up all the demand volume originating from and destining to that hub region and dividing by doubled total demand volume (see Eq.6-7). The five major hub regions cover 43.55% of the total demand volume, which means the other 64 hub regions cover only 56.45% of the total demand volume. The index “Involved self-owned aircraft” counts the number of self-owned aircraft originating from and destining to the hub. With 27 self-owned aircraft, 2/3 [36/ (27*2)] of the self-operated flights either originates from or destines to the five hubs. Moreover, it seems that Hub 23 serves as a major air hub in the air network, while Hub 41 serves as a regional air hub. Just as the conclusion of Hub Arc Problem by Campbell et al[450], the model presents automatically a quasi H/S air network although we do not designate the structure of the air network.

$$\text{Demand volume covered by hub } k\ (\%) = \frac{\sum_{i\in N}\sum_{j\in N}(w_{ij}+w_{ji})x_{ik}}{\sum_{i\in N}\sum_{j\in N}(w_{ij}+w_{ji})}\times 100\% \qquad k = 9, 23, 41, 49, 169 \tag{6-7}$$

[449] That is the Economic Circle over the Yangtze River Delta (hub region #23), the Economic Circle over the Pearl River Delta (#41), the Economic Circle over Bohai Bay (#9), the Economic Circle of Chenyu (#49) and the Economic Circle over Yellow River Delta (#169).

[450] See Campbell et al (2003), pp.555-574; Campbell et al (2005a), pp.1540-1555; Campbell et al (2005b), pp.1556-1571.

- **Aircraft fleet update strategy**

Nr.	Fixed cost k_p (/km)	Variable cost χ_p (/kilo*km)	Lower bound u_p^l (kilo)	Upper bound u_p^u (kilo)	Current	Optimal
1	0	0.007	0	∞	4665	4455
2	0.3	0.0068	0	2700	0	193
3	1.5	0.0065	0	5000	2	0
4	3.5	0.006	0	8000	10	44
5	14	0.0055	0	15000	15	0

Table 6-12: Comparison between current aircraft fleet and the optimal fleet composition with *Ext.1*

When we compare the current aircraft fleet composition with the optimal fleet composition resulting from *Ext.1* (see the last columns in Tab.6-12), we discover that the current fleet is quite unreasonable to the newly-planned network.

In our case Service type 2 represents a kind of light cargo plane in the fleet with relatively small fixed cost and capacity but high variable cost, while Service type 5 represents a kind of large cargo plane in the fleet with relatively low variable cost but large fixed cost and capacity. Although large aircraft are economical to the current business pattern and network structure, i.e. large demand volume among few hubs, they are not suitable to the new network, in which small inter-hub flows occur among much more hubs. For this reason, large aircraft are not adopted by the new network, while small aircraft that are not included in the current fleet are in demand. Meanwhile, we can also observe that although Service type 3 is smaller than Service type 4 with smaller capacity, higher variable cost and lower fixed cost, Service type 3 is not included in the optimal fleet. A second check of the air cost functions reveals that Service type 2 is always more economical than Service type 3.

Company A also needs an aircraft fleet update strategy to guide the implementation of the project. Subject to the budget and other management constraints, the company could only update the aircraft fleet composition step by step, suppose by each budget period. The dynamic fleet update decision is made based on the latest information about demand and cost. Although the total cost of the network during the updating period will be higher if the aircraft fleet is updated to optimization step by step rather than all at once, the advantage of the former is also attractive. The decision risk is much lower with less uncertainty about demand and cost. It is therefore an effective measure to evade decision risks from new services in emerging markets.

Since the hub location decision is not sensitive to air service constraints, at the end of each budget cycle the H/S system can be updated to the optimum with the best choice among the available aircraft fleet compositions at that time by changing air service selection decisions and demand allocation decisions. We illustrate this decision process with an example, which is based on the following three assumptions.

(1) Demand node can be changed from one "home" hub to another without cost[451].

(2) The budget for each period can afford, for example, two aircraft of Service type 4 or three aircraft of Service type 2.

(3) Only one unreasonable aircraft in the current fleet can be disposed during each decision period if necessary, subject to some management constraints. The disposal cost is ignored.

We take the current state as "Phase 0" and make the update decision for "Phase 1" with the help of the decision tree in Fig.6-13 ("S" denotes service and "C" denotes constraint). In Phase 0, the total cost is 56540386 under the constraints of current available aircraft. Then we calculate the total cost with fleet composition of 2 additional Service type 4 or 3 additional Service type2 in combination of 1 less Service type 5 or 1 less Service type3, respectively. The aircraft fleet is updated with the fleet compositions with the least total cost (mark with shadow in Fig.6-13). This process can be repeated at each budget period even when the aircraft fleet composition conforms to the optimal composition with *Ext.1*, since the demand and cost are always changing.

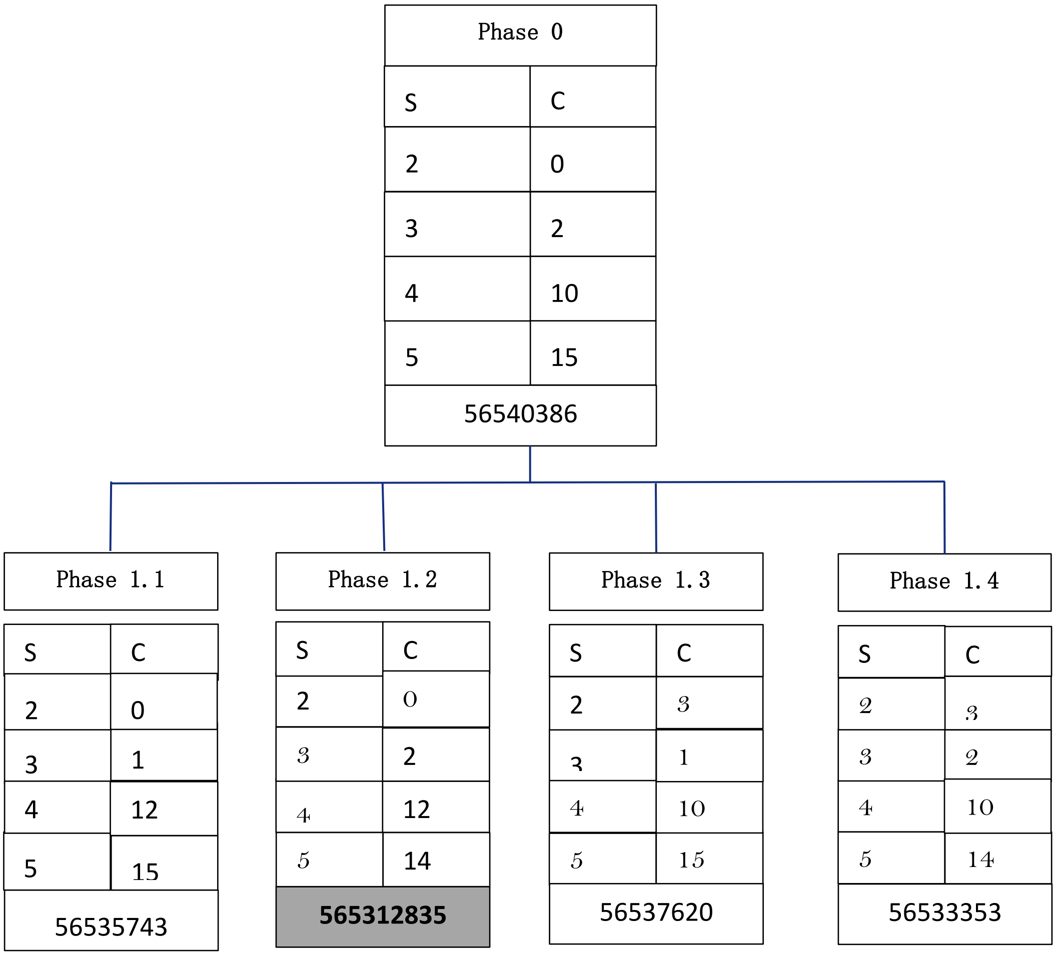

Figure 6-13: Decision tree for aircraft fleet updating

[451] But in our case it is not true. Company A regards the decisions concerning the network structure, i.e. hub location and demand allocation as long-term decisions, while air service selection and feeder routing are medium-term decisions. Because regional daughter companies are established on the basis of the network structure, changes of demand allocation mean changes of customer resource, assets and also share-holding between daughter companies. On the contrary, air service selection is only internal decision of airline subsidiary company.

6.3. Empirical scenario planning based on *Ext.2*

A well designed network means it does not only perform well under current system state but also continue to be profitable for the system's lifetime and stand up when environmental factors change or market trends evolve. Robust network planning is thus a difficult task, demanding that decision makers account for uncertain future events. In this section, we make scenario planning to help decision-makers capture uncertainty and the corresponding decision risk by specifying a number of possible future scenarios[452]. Planners consider some strategic "what-if" assumptions quantitatively and see how the model will react if the values of input data derive from the present state[453].

Specifically speaking, scenario planning is carried out here with two purposes. One purpose is to find those factors that are critical to the final decision by checking the discrepancies of the solutions under different scenarios. These factors are divided into two types, i.e. controllable and uncontrollable (see Fig.6-14). Uncontrollable factors, such as hub fixed cost and demand volume, are actually risk that Company A must be faced up with. Those critical factors must be estimated as carefully as possible. Controllable factors, such as loading factor and hub coverage radius, are actually network policy set by network planners. These policies must be seriously considered if the network decisions, especially those long-term decisions, are sensitive to them. Another purpose of the scenario planning is to provide us several different backup solutions with near minimum cost. For one thing, decision-makers sometimes refuse to adopt solution with the minimum cost due to some managerial factors or other considerations. For another, the solution with minimum cost is not guaranteed to be optimal.

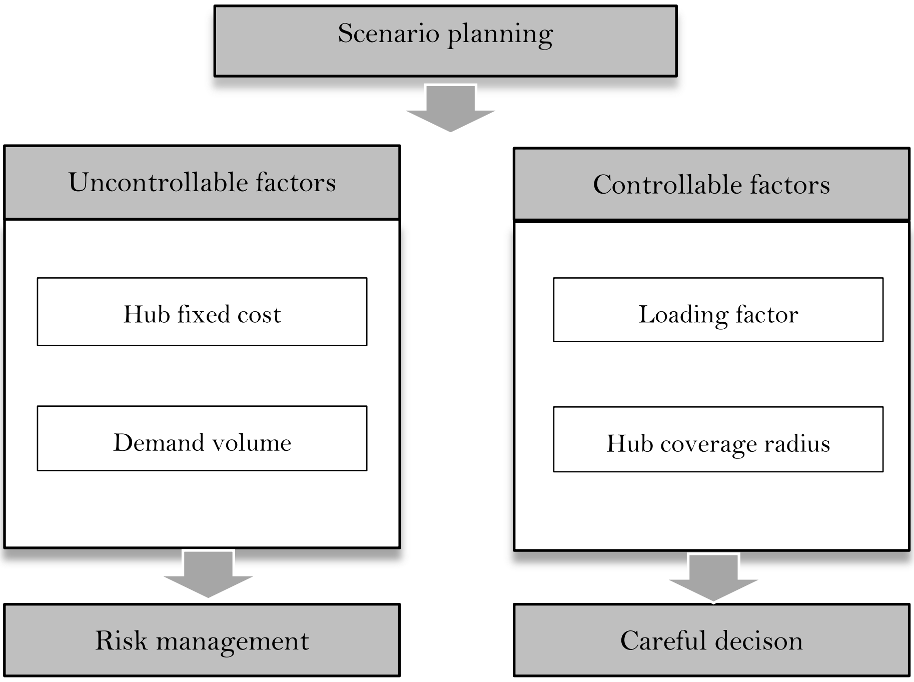

Figure 6-14: Purposes and elements of scenario planning

[452] See Mobasheri et al. (1989), pp. 31-44.

[453] See Mulvey (1996), pp. 1-15.

All the computational tests for scenario planning in this section are based on current situation (i.e. *Ext.2*) and run with the same algorithm parameters as those in Sec.6.2. Running time of the hybrid GAs is less than 1900 seconds under all instances. We continue to base our analysis on the best-known solution after 10 runs for each instance. The CV values under all instances are less than 0.0025. Moreover, we also make *T*-tests between every two instances with the 10 best solutions for each instance to make sure that the differences between the analyzed solutions are significant, i.e. those differences between solutions result from the variations of the control parameters rather than from the randomness or instability of the algorithm, indicating our analysis is reliable and creditable. When the result of *T*-test shows that the difference between the two groups is not significant, we increase the variation of the control parameter. The insignificance results from the instability of the hybrid GAs itself.

6.3.1. Variation of hub fixed cost fh_k

The hub fixed cost is one of the factors with large uncertainty in the model due to the high fluctuation of the real estate price in China. The real estate price not only depends on national economic situation but also is largely controlled by the central government. Meanwhile, local government may have some special policies in some regions. Under these considerations, we not only neutrally estimate the hub fixed cost node by node but also create several scenarios that vary from neutrality. Tab.6-13 presents these scenarios and the corresponding solutions with *Ext.2*. The variations of the hub fixed cost are estimated both negatively and positively with a step length of 10% from the neutral scenario. Since the expectation on price increase is much higher than that on price decrease, we also include the scenario a positive increase of 30%.

Variation level	Hub fixed cost	Hub Nr.	Air cost	Feeder cost	Total cost
-20.00%	27600000	69	19555555.15	2484830.55	49640385.7
-10.00%	31050000	69	19555555.15	2484830.55	53090385.7
0.00%	34500000	69	19555555.15	2484830.55	56540385.7
10.00%	37950000	69	19555555.15	2484830.55	59990385.7
20.00%	41400000	69	19555555.15	2484830.55	63440385.7
30.00%	44850000	69	19555555.15	2484830.55	66890385.7

Table 6-13: Scenario planning for variation of hub fixed cost

The best-known solutions under each instance in Tab.6-13 indicate that the hub number, hub location and even the demand allocation decisions remain the same with the variations of the hub fixed cost between (-20%, 30%). The network configuration appears to be not sensitive to the hub fixed cost. The variation of the total cost comes totally from the variation of the hub fixed cost. As we have mentioned in Sec.6.2.3, the hub fixed cost is probably still so large even with a -30% variation that the model always choose as few hubs as possible to minimize the total cost.

However, the result of the tests can only be applied to the case, in which the hub fixed cost is the same. In real life instance, the hub fixed cost is estimated one by one based on “sunk cost theory”. For this reason, the situation in real-world must be much more complicated than the test instances here.

6.3.2. Variation of demand volume w_{ij}

Geographically-uneven economic development in China results in uneven distribution of EDS demand. After the first round development of large cities along the east coast, the second round development is approaching to large cities in west of China and second tier cities in east of China. For this reason, we believe the fluctuation degree of the demand varies in different types of cities.

We divide all the 281 cities in the demand nodes set N into two groups based on the result of "market segment" in Sec.6.1.1 (See Tab.6-2). We regard the 92 cities in Group 1 and 2 as key cities and the 189 cities in Group 3 as normal cities. It is believed that key cities are economically developed ones with more stable demand volume than normal cities. Thereby, the demand volume between normal city and normal city may comparatively have the largest fluctuation, followed by that between normal city and key city, and by that between key cities. Three scenarios, namely negative, positive and explosive to the neutral one, are developed (see Tab.6-14). The corresponding solutions, together with that of the neutral scenario, are illustrated in Tab.6-15. The values in the parenthesis represent the corresponding percentages based on the neutral scenario.

Scenarios	Negative	Neutral	Positive	Explosive
key/key city	-10%	0	10%	20%
key/normal city	-15%	0	15%	25%
normal/normal city	-20%	0	20%	30%

Table 6-14: Demand volume deviation under different scenarios

Scenarios	Negative	Neutral	Positive	Explosive
Hub Nr.	69	69	69	69
Total cost	53236867 (94.16%)	56540386	59789303 (105.75%)	60878999 (107.67%)
Total air cost	16591242 (84.84%)	19555555	22386404 (114.48%)	23349887 (119.40%)
Cost by self-owned aircraft	1075051 (67.77%)	1586280	2007222 (126.54%)	2122665 (133.81%)
Cost from air freight market	15516194 (86.35%)	17969275	20379182 (113.41%)	21227221 (118.13%)
Hub fixed cost	34500000 (100%)	34500000	34500000 (100%)	34500000 (100%)
Feeder transportation cost	2145621 (86.35%)	2484830.55	2902899 (116.82%)	3029112 (121.90%)
Volume by self-owned aircraft (kilo*kilometer)	158938703 (67.09%)	236898338	302651297 (127.76%)	321240312 (135.60%)
Volume by air freight market (kilo*kilometer)	2216599150 (86.35%)	2567039273	2911311679 (113.41%)	3032460186 (118.13%)
Volume by truck (kilo*kilometer)	429124200 (86.35%)	496966110	580579800 (116.82%)	605822400 (121.90%)

Table 6-15: Scenario planning for variation of demand volume

The results under the four scenarios show that the variation of the demand volume has no impact on the hub number within the fluctuation range under consideration. But actually it has impact on the hub locations, which cannot be reflected in this table. 3 hubs in the explosive scenario have changed the location compared to the other three scenarios. It can be inferred that although the uneven variations of the demand volume may result in the change of the hub locations, the model still tries to choose as few hubs as possible to minimize the total cost.

EOS is clearly reflected with the values of "total cost" under the four scenarios. The deviations of the "total cost" from the neutral scenario are -5.84%, 5.75% and 7.67% for the negative, positive and explosive scenarios respectively, while the demand volume deviations of these three scenarios are at least -10%, 10% and 20%. The EOS comes primarily from the hub fixed cost, since the hub number in these three scenarios keeps the same. It also comes from the self-owned aircraft. Observe that the deviations from the neutral scenario of "volume by self-owned aircraft" and "cost by self-owned aircraft" are larger than the deviations of "total air volume" and "total air cost" (see Tab.6-16).

	Negative	**Neutral**	**Positive**	**Explosive**
Cost by self-owned aircraft	-32.23%	0.00%	26.54%	33.81%
Total air cost	-15.16%	0.00%	14.48%	19.40%
Volume by self-owned aircraft	-32.91%	0.00%	27.76%	35.60%
Total air traffic volume	-15.28%	0.00%	14.62%	19.61%

Table 6-16: Deviation of the air cost and air traffic volume from demand variations

6.3.3. Variation of loading factor

The loading factor is a controllable factor for decision-makers to balance between cost and risk. A high loading factor means high risk that demand volume in peak season is beyond the capacity of self-owned aircraft so that the rest demand must be fulfilled by commercial air service separately. However, self-owned aircraft are better loaded in normal season with a high loading factor. The loading factor under the basic instance is set to be 80%. In this section, we increase the loading factor from 60% to 90% with a step length of 10% and test how it can affect decisions and corresponding costs. The capacity upper bounds of each type of self-owned aircraft with different loading factors are listed in Tab.6-17.

Upper bound u_p^u (kilo)	**60%**	**70%**	**80%**	**90%**
Service type 1	∞	∞	∞	∞
Service type 3	3750	4375	5000	5625
Service type 4	6000	7000	8000	9000
Service type 5	11250	13125	15000	16875

Table 6-17: Capacity upper bound of self-owned aircraft with different loading factors

Loading factor	**60%**	**70%**	**80%**	**90%**
Hub Nr.	69	69	69	69
Total cost	56541046	56540897	565403865	56538018
Total air cost	19556216	19556067	19555555	19553184
Cost by self-owned aircraft	1173847	1364201	1586280	1729045
Cost from air freight market	18392139	18193144	17969275	17816713
Hub fixed cost	34500000	34500000	34500000	34500000
Feeder transportation cost	2484829	2484830	2484831	2484834
Volume by self-owned aircraft (kilo*kilometer)	176489262	204917063	236898338	258692985
Volume by air freight market (kilo*kilometer)	2627448362	2599020583	2567039273	2545244725
Volume by truck (kilo*kilometer)	496965881	496965997	496966110	496966834

Table 6-18: Scenario planning for variation of loading factor

The best-known solutions under each scenario are displayed in Tab.6-18. However, the results of T-tests indicate that the differences between the scenario 60% and 70%, 70% and 80%, and 80% and 90% are not significant. For this reason, any analysis based on the neighboring scenarios is not reliable. However, the overall trend of the solutions is reliable, since the differences between the scenario 60% and 80%, 70% and 90%, and 60% and 90% are significant.

The solutions indicate that the hub location decision is not sensitive to the loading factor with the variations between (60%, 90%). Discrepancies of total cost result from small changes on allocation decision and air service selection decisions on some links. With the increase of the loading factor, the volume by self-owned aircraft increases so that the air cost, as well as the total cost, decreases. However, the variations between the solutions are very little due to the limited number of self-owned aircraft.

However, this analysis only involves key EDS of the company (see Tab.1-2). As we have mentioned above, the spare loading space on board of self-own aircraft is in practice filled with other economical express demand. The cost discrepancies between these scenarios are, therefore, not as large as those shown in Tab.6-18 at the company level.

6.3.4. Variation of hub coverage radius D

In this section, we examine how the variation of the hub coverage radius impacts on the solution of the model. Three hub radii listed in Tab.6-19 are under consideration. The corresponding time windows for tributary network and backbone network are also listed below based on an average highway speed of 90km/h and the time schedule in Tab. 6-3. We also assume direct flight between any potential hubs is less than 5 hours.

Hub coverage radius (km)	**225**	**270**	**315**
Time window for tributary network (hour)	2.5	3	3.5
Time window for air network (hour)	7	6	5

Table 6-19: Hub coverage radius and corresponding time window for tributary and backbone networks

Hub coverage radius	**225**	**270**	**315**
Hub Nr.	94	69	55
Total cost	68713053 (121.53%)	56540386	49539818 (87.62%)
Total air cost	19639177 (100.43%)	19555555	19303250 (98.71%)
Cost by self-owned aircraft	919735 (57.98%)	1586280	2280160 (143.74%)
Cost from air freight market	18719442 (104.17%)	17969275	17023090 (94.73%)
Hub fixed cost	47000000 (136.23%)	34500000	27500000 (79.71%)
Feeder transportation cost	2073876 (83.46%)	2484831	2736568 (110.13%)
Volume by self-owned aircraft (kilo*kilometer)	136427553 (57.59%)	236898338	345386351 (145.80%)
Volume by air freight market (kilo*kilometer)	2674205929 (104.17%)	2567039273	2431870050 (94.73%)
Volume by truck (kilo*kilometer)	414775200 (83.46%)	496966110	547313600 (110.13%)

Table 6-20: Scenario planning for variation of hub coverage radius

The solutions under the three scenarios are tabulated in Tab.6-20. Note that the hub location decision is quite sensitive to the controllable factor hub coverage radius. When the hub coverage radius decreases, the number of hubs in the network, the total cost and the transportation cost increase. Although the feeder transportation cost decreases, it cannot compensate the increase of the air cost.

When we go through the traffic volume, we find that the reduction of the hub coverage radius leads to the decrease of volume by truck but the increase of volume by air. However, the average traffic volume on inter-hub links and thus the traffic volume by self-owned aircraft decrease.

A small hub coverage radius means short delivery time if the time window for the air backbone network is fixed. But it also means high total cost that results from the increase of the hub number and air traffic volume. In fact, these two measures of network performance, i.e. delivery time and cost, are conflictive and any gain in one is expected to be accompanied by a loss in the other. For example, with the assumption of a 5-hour time window for the air network the delivery time decreases from 12 hours to 10 hours by 16.67%, when the hub coverage radius decreases from 315km to 225km. However, the reduction of the delivery time leads to a somewhat steady increase of the total cost by 38.7%. If we interpret the delivery time as an indicator of the service quality, we may suggest that a smart company may only allow moderate deterioration of the service quality which leads to steady cost reduction. Otherwise, the mild cost saving cannot compensate the sharp reduction of the revenue.

When we go through all the hub location decisions under the four scenarios, we find that some hubs always stand in optimal solutions. The first kind of these hubs, taking Urumqi as an example, is geographically dispersed from other potential hubs. Actually, they may be designated as hubs in preprocess procedure. The second kind is hubs with large origin or destination flow themselves, such as Peking and Shanghai. The third kind is hubs with location advantage. Although their in-and-out demand volumes are not so large, they locate at the center of node clusters. These hub locations are not sensitive to the other factors so that they can be built with priority if the budget is limited.

However, three concerns must be further considered in real life. First, the model is based on the assumption that hubs are fully inter-connected. Once stopover is allowed, a loose time window for the air network means large transfer opportunities and thus air cost saving. Second, the model assumes that demand nodes are directly connected to "home" hubs. However, pure star-shaped feeder networks seldom appear in reality. When demand nodes in the hub regions are connected by several routes rather than by direct service, small hub coverage radius also means less feeder trucks and thus cost saving of feeder transportation. Third, the service quality of the same day EDS within the hub region is not considered here. Although these three concerns are not included in this dissertation, they must be considered in reality.

6.4. Summary

This chapter is devoted to empirical study on real-life problem. In Sec.6.1 we illustrate how we collect and modify input data of the models. We illustrate the problems we are faced up with and introduce the methods and mathematical instruments we have used. The purpose of this section is to provide our readers an overview of the project and some guidance to the application of the proposed models and algorithms.

In Sec.6.2 we display the solutions of *Ext.1* and *Ext.2* by the proposed hybrid GAs under the basic instance. The comparison between them not only suggests some important features of our specific network but also indicates some general conclusions: (1)the cost focus shifts from the tributary networks in pure ground H/S networks to the backbone air network in air-ground H/S networks, which indicates that the planning focus of multimodal EDS networks should also lie in the air network; (2) the concave piecewise linear cost function (that can be easily transformed from the cost selection function in this dissertation) has flow bundling effect; (3) models with concave piecewise linear cost function may automatically present a quasi H/S network although no such structure is imposed. In this section we also provide a dynamic update strategy of aircraft fleet to guide the implementation of the project. We test the performance of the Improvement technique 5 in Sec.4.2.5 with the project data set. However, its time-saving advantage is not significant in our case probability due to the short calculation time of the embedded integer programming and the fluctuation of the GAs' running time.

Sec.6.3 is devoted to scenario planning based on *Ext.2*. Most of the results in scenario planning suggest that the hub fixed cost in neutral scenario is relatively high so that the model always chooses as few hubs as possible to minimize the total cost. For this reason, the hub location decision is not sensitive under most scenarios, except the hub coverage radius. The hub coverage radius seems to be decisive to the network configuration in our case and must be set with great care. Some general conclusions are obtained or verified: (1) EOS can be obtained in H/S network, which primarily comes from the hub fixed cost and may also from self-owned aircraft; (2) loading factor is a controllable factor for decision-maker to balance costs and the corresponding over-demand risk;(3) the hub coverage radius is a controllable factor to determine the service quality in terms of delivery time according to the cost. However, scenario planning in this section is under some simplifications of the reality which must be considered in real-life cases.

7. Conclusion and perspective

7.1. Summary of research and contributions

This dissertation focuses on strategic planning of large-scale, multi-modal and time-definite EDS networks, while trying to provide readers with an overview of the project.

In order to provide trans-city overnight EDS among relatively developed cities in China, we resort to an air-ground H/S network with fully interconnected/star shaped structure. The corresponding models are a combination of Hub Location Problem with Fixed Cost[454] and Hub Set Covering Problem[455]. Because neither of the two problems simultaneously addresses both of the two issues (1) the time constraints and (2) the total cost, especially transportation cost. We combine these two problems together to complement each other. We minimize the total cost (both hub fixed cost and transportation cost) under the constraints that all demand nodes are within the coverage of their "home" hub. We first propose the basic model, which conforms to the three conventional assumptions, i.e. fully interconnected hubs, fixed discount rate on hub arcs and no direct link between non-hub nodes. Then we extend the basic model by eliminating the assumption of fixed discount rate on hub arcs. Air service selection problem is embedded for the backbone air network based on a cost select function that can be easily transformed into piecewise linear function. We consider two different situations-whether the air service selection is subject to the current fleet composition (*Ext.2*) or not (*Ext.1*).

Due to the large scale of our real-life instance, we resort to hybrid GAs to get good solutions in bearable time but without the guarantee of finding optimal solutions. In particular, the overall problem has three kinds of decisions, i.e. hub location decisions, demand allocation decisions and air service selection decisions. We propose for each kind of decisions one specific algorithm, namely, GAs, local search heuristics and integer programming, respectively. These three algorithms run successively in a hierarchical framework to solve the original problem. In order to improve the performance of the algorithm, we propose 5 improvement techniques, which are applied to different procedures of the original algorithm.

Computational tests based on public data set and project data set are conducted to evaluate the performance of the proposed algorithm in terms of computational time and solution quality. Tests under small-scale instances with CAB data set are to evaluate the overall performance of the proposed algorithm by comparing the solutions with the optimal solutions generated by CPLEX. The results indicate that CPLEX (exact method) generally outperforms the hybrid GAs both in terms of running time and solution quality under instances with no more than 25 nodes. Tests under large-scale instances with AP data set and project data set are to evaluate the performance of the 5 proposed improvement techniques. Since neither the optimal solutions nor solutions generated by other algorithms under large-scale instances are available to serve as benchmarks, we provide information about the stability of the solutions with CV values and the reliability of the results with T-tests. The 5 improvement techniques and their performance with test data are briefly summarized as follows.

---

[454] See e.g. Cunha/Silva (2007), p.750; Chen (2007), p.213.In some literature, the union of uncapacitated hub location problem (UHLP) and capacitated hub location problem (CHLP) is actually hub location problem with fixed cost in this dissertation. See e.g. Hekmatfar/ Pishvaee (2009), pp. 243-270.

[455] See e.g. Alumur/Kara (2008), pp.9-11 and p.14.

Improvement 1 refers to the constructive procedure for the initial solution generation procedure of GAs. We incorporate the constructive procedure borrowed from GRASP to generate initial solutions for GAs. This method can not only yield feasible solutions but also balance the diversity and intensity of the initial solution pool. Test results prove that high quality solutions and large diversity are desirable characteristics of initial solution pool for GAs. However, a good parameter and a large solution space are prerequisites for the positive effect of constructive procedure on GAs.

Improvement 2 refers to the injection mechanism for the update of solution pools after each generation of GAs. The test results indicate that the injection mechanism can improve the solution quality of GAs by increasing the diversity of the solution pool during the process of GAs with a moderate injection rate. High injection rates may break up the balance between diversity and intensity, and lead GAs to randomness. Moreover, the impact on the GAs performance from the injection mechanism and constructive procedure in Improvement 1 is relatively independent and can hardly compensate to each other.

Improvement 3 refers to LS after GAs on both hub location decisions and demand allocation decisions. The test results indicate that the time-consuming full LS on hub location decisions does not work so efficiently after GAs with regard to the computational time and potential improvement. Meanwhile, full LS on allocation decisions can time-efficiently improve the best solution by the hybrid GAs. The test results also consolidates the idea to apply partial LS on allocation decisions during the GAs, since hub location decisions play a decisive role in determining the solution quality of GAs. It is not worth much effort to improve the demand allocation decisions until the hub location decisions are good enough.

Improvement 4 refers to a "good" initial solution for LS on allocation decisions. The test results indicate the demand allocation pattern of "minimum cost" performs the best, followed by the pattern of "maximum flow", and then by the patterns of "multi-criterion" and "nearest-distance". The test results indicate that for air-ground H/S systems, in which the backbone cost rate is higher than the feeder cost rate and demand nodes are unevenly distributed, the "minimum cost" allocation pattern is more suitable, although the running time is a little bit longer.

Improvement 5 refers to an approximation of the sub-problem solution in the early stage of GAs to improve the time-efficiency of the algorithm. It is exclusive for Ext.2, in which integer programming is necessary for the air service selection decisions. Generally speaking, if the algorithm for the sub-problem is time costly and must be invoked frequently, this time-saving method can be efficient as long as the approximation is relatively good. However, the time-saving effect is not significant in our case may due to the short calculation time of the embedded integer programming and fluctuation of GAs' running time.

Finally, the models and the tailored hybrid GAs are applied to real-life instances of the project. We introduce how we collect and modify the input data and how we deal with problems we are faced up with. By analyzing and comparing the solutions of *Ext.1* and *Ext.2* under the basic instance, we not only reveal some important features of the network but also get some general conclusions and provide a dynamic aircraft fleet update strategy to guide the implementation of the project. Finally, scenario planning is executed to help decision-makers identify critical uncontrollable and controllable factors to balance between costs and corresponding decision risks.

This dissertation is supposed to make the following advances and contributions.

1. Compared to other works on the strategic network planning, our models are applicable to rebuilding or modifying the current network rather than building a brand new network.
2. We make a relatively comprehensive review on HLPs and make innovative comparisons between HLPs and FLPs from various aspects.
3. We propose a generalized cost select function for different service types and smooth the unreasonable gap in the formulation.
4. We propose 5 improvement techniques for different procedures of GAs. Computational tests show that the performance of the tailored hybrid GAs is significantly better than the simple untailored one.
5. Our research is application-oriented. We present our readers an overview of the project, from enterprise strategy interpretation, service definition, model data collection and modification to project implementation strategy.
6. Air-ground EDS network in our dissertation differs from those pure ground H/S networks in terms of time constraints and cost distribution. We reveal some general features of the air-ground H/S network and also suggestions for the corresponding heuristics.

7.2. Limitations and future research

Despite the above-mentioned advances and contributions, this dissertation has several limitations remaining for further research.

- Effective meta-heuristics for large-scale network planning problems.

Although computational tests with the AP data set indicate that the proposed improvement techniques can improve the performance of the algorithm significantly compared with the untailored GAs, a fundamental problem remains in this dissertation. We have neither optimal solutions nor solutions by other algorithms that can serve as benchmarks to evaluate the performance of the proposed algorithm under large-scale instance. Future researches may solve this problem optimally with exact algorithm or develop other heuristics taking ours as a benchmark.

- Interdisciplinary research about network planning

As a sub-topic of OR, network planning always considers factors only concerning the network itself, such as costs, transportation time, consolidation centers or hubs, vehicles and routings. However, from the point view of the owner or the manager of the network, service quality (such as delivery time and punctuality rate), cost and revenue are correlated. In other words, the result of the network planning can be better when the principles of marketing, revenue management and even financial management are also under consideration. Interdisciplinary research is in demand in practice. However, up till now, only few papers expand network planning models by considering factors, such as pricing[456], demand management[457] and market competition[458].

[456] See e.g. Yan et al. (1995), pp.171-180.

[457] See e.g. Dasci/ Laporte (2005), pp.397-405.

[458] See e.g. Colome. et al. (2003), pp. 121-139; Drenzner et al., (2002), pp. 138-151; Lin/ Lee (2010), pp.618-629; Gelareh/ Nichel (2010), pp.991-1004; Martin/ Roman (2004), pp.135-150. Sinha (2004), p. 51-61.

Just as overwhelming majority of researches on network planning, our model principally gives up the ultimate goal of enterprises, i.e. profit, by ignoring service pricing, market competition and relationship between service quality and system cost. We take fixed demand as the model input and “cost minimization” as the objective. Future research may take “profit maximization” as the objective by simultaneously considering service pricing problems or competition factors.

- HLRPs with backbone network

Owing to the large scale of the real-life instances, the network planning problem in this project is divided into several sub-problems. This dissertation focuses on strategic planning. However, the dividing of the overall planning problem results in sub-optimization of the system. Oversimplified assumptions on feeder and backbone network configuration can seriously distort route length, transportation time and corresponding cost.

Compound network planning problems (also called combinatorial problems) that simultaneously include facility location and vehicle routing problems, come into being for this reason. As we have mentioned in Sec.2.2.2, HLRPs have received more and more attention. Just as LRPs, HLRPs determine hub location and feeder routing simultaneously with the assumption that hubs are fully interconnected. This is reasonable in pure ground H/S network since there is always cost discount for backbone link due to EOS. However, such assumption is not suitable for EDS network planning. As we have concluded in Sec6.2.3, feeder networks in air-ground H/S systems are less important than those in pure ground H/S systems both in terms of delivery time and cost. With the requirements of shorter service time and longer delivery distance, the covered area by each ground hub becomes smaller, while the backbone air network plays a decisive role in determining the total delivery time and cost. Therefore, it is necessary to include backbone network planning in network planning problems for EDS.

In this dissertation, we include a cost selection function for different service types in the hub network based on the assumption that all hubs are fully interconnected by direct flights. Future research may include aircraft routing problem with constraints of one or two stopovers. Advances in mathematical programming methodology and improvements in computer technology are likely to enable researchers to effectively solve even more difficult models in the future.

- Hierarchical HLPs

Hierarchical HLPs consider more than one type of hubs in the models. Compared with hierarchical FLPs[459], hierarchical HLPs are more complicated for hubs are connected. There are only few papers involve this topic[460].

Results of our extension models indicate that a quasi H/S air network comes into shape although we do not impose the structure of the air network in the models (see Sec.6.2.3). So it is interesting to model our case as a hierarchical HLP and compare its result with ours.

- Network planning for multiservice

---

459 See e.g.Costa et al (2011), pp.3-13. Review on this topic, please refer to Sahin/ Süral (2007), pp. 2310-2331.

460 See e.g.Yanman (2009), pp. 643-658; Lin /Chen (2008), pp.986-2003; Lin (2010), pp. 20-30.

In practice, the network planned in this project also supports economical EDS and logistics service, although we do not include them in this project. Integrated network is a cost-saving strategy that is commonly adopted by EDS providers. Perhaps the best example is United Parcel Service (UPS), which offers both overnight packet service and deferred delivery service domestically with an integrated air-ground network. Priority for sorting and dispatching is naturally given to packet delivery service. However, as the cost of transporting deferred packets by air is marginal, if excess capacity exists, some deferred delivery orders are also dispatched by air. The advantages of such integration include increased customer density, flexible modal choice and reduced cost in ground transportation system.

Network planning for multiservice is initialized by Smilowitz and Daganzo[461]. The corresponding location and routing problems must comply with various time constraints for all service types under consideration. For deferred packet service, time constraints are somewhat relaxed and more cost-efficient routings are possible. However, with the number of routing options and service types increasing, finding an optimal network configuration becomes more difficult.

- Network planning under uncertainty

Making robust long-term decisions for network planning is a tough task, requiring decision-makers to account for uncertain events in the future. The complexity of HLPs has limited most of the studies to static and deterministic models. So are our models, although we carry out scenario planning to consider uncertainty in the future. However, the scenario approach has two main drawbacks, although it is more tractable. One is that identifying scenarios is a daunting and difficult task; indeed, it is the focus of a large body of stochastic programming literature. The second disadvantage is that only a relatively small number of scenarios are considered for computational reasons[462].

It is therefore attractive to apply stochastic optimization, robust optimization or dynamic optimization to consider uncertainty in the future. Stochastic optimization attempts to capture the uncertainty in input parameters by considering the probability distribution of uncertain parameters[463]. Robust optimization attempts to optimize the worst-case performance of the system, if there is no information about the probabilities of the parameters[464].Dynamic optimization focus on the timing issues involved in network over an extension horizon. Some current researches on HLPs under uncertainty[465] may serve as the starting of future research.

[461] See Smilowitz /Daganzo (2004), pp.4-6; Smilowitz /Daganzo (2007), pp.183-196.

[462] See Snyder (2006), p.4.

[463] See e.g. Santoso et al., (2005), pp. 96-115; Sim et al. (2009), pp.3166-3177. For literature review, please refer to Snyder (2006), pp.547-564 and Owen/ daskin (1998), pp.423-447.

[464] See Snyder (2006), p.3.

[465] Synder listed some in a literature review. See Snyder(2006), pp. 547-564.Recent papers on such topics include e.g. Sim et al (2009), pp.3166-3177.

Appendices

Appendix 1: Potential hub set and corresponding fixed cost

No.	Node in AP	total weight	fixed cost
1	159	58391053	20369.65
2	151	26503445	29955.73
3	27	11779979	22009.12
4	160	11258636	33061.82
5	161	10867159	30827.59
6	42	10514268	26028.58
7	147	9811386	31839.25
8	157	8925085	27440.08
9	101	7296215	21447.45
10	129	6858623	27293.22
11	19	4118603	25424.19
12	156	4060176	25497.44
13	35	4031481	27053.44
14	66	3819714	33029.2
15	20	3769378	25355.03
16	59	3655650	31007.65
17	174	3651760	21231.27
18	165	3648752	25441.26
19	131	3544529	25686.59
20	150	3519698	20449.16
21	16	3495011	20860.35
22	170	3428563	22671.23
23	158	3339916	31469.84
24	127	3244088	24441.51
25	92	3208164	28813.37
26	163	3124065	22456.05
27	107	3076706	20753.08
28	67	3058345	23585.65
29	39	3038405	30763.81
30	33	3034958	23013.29
31	189	3002914	23526.63
32	187	2995127	32862.29
33	166	2958361	33245.54
34	137	2801588	33385.3
35	58	2669423	34616.57
36	142	2648898	32275.04
37	1	2639783	30964.59
38	139	2570943	33532.03
39	65	2543094	28159.61
40	70	2494373	32626.96

41	93	2456174	20055.49
42	34	2421298	22194.28
43	154	2272046	29430.92
44	183	2121545	28914.75
45	7	2079159	26429.1
46	125	2056852	32875.89
47	140	2036597	30975.55
48	71	1987691	22064.61
49	72	1921696	32001.95
50	146	1911556	30807.07
51	37	1909311	20369.65
52	149	1901094	29955.73
53	134	1865224	22009.12
54	135	1852397	33061.82
55	41	1845927	30827.59
56	63	1773928	26028.58
57	172	1727741	31839.25
58	167	1727653	27440.08
59	106	1715152	21447.45
60	100	1671061	27293.22
61	196	1666190	25424.19
62	81	1661049	25497.44
63	175	1645545	27053.44
64	24	1642800	33029.2
65	112	1617662	25355.03
66	192	1605770	31007.65
67	132	1584104	21231.27
68	48	1578746	25441.26
69	171	1576818	25686.59
70	79	1535694	20449.16
71	141	1518471	20860.35
72	6	1471364	22671.23
73	84	1461606	31469.84
74	44	1449914	24441.51
75	49	1432283	28813.37
76	62	1386590	22456.05
77	133	1385187	20753.08
78	3	1354192	23585.65
79	4	1348103	30763.81
80	117	1301131	23013.29
81	177	1252730	23526.63
82	162	1252605	32862.29
83	75	1220268	33245.54
84	78	1169141	33385.3
85	176	1167777	34616.57
86	148	1159292	32275.04
87	199	1116667	30964.59

88	10	1115022	33532.03
89	185	1106757	28159.61
90	110	1105288	32626.96
91	152	1102605	20055.49
92	89	1079243	22194.28
93	80	1074254	29430.92
94	98	1058389	28914.75
95	43	1054869	26429.1
96	180	1045760	32875.89
97	111	1040118	30975.55
98	190	1039124	22064.61
99	102	1026202	32001.95
100	45	997274	30807.07

Appendix 2: Results of independent two-sample *T*-tests on reproduction parameter setting

Test- 1: Best solutions under $(H=50, P_{cro}=0.7, P_{mut}=0.2)$ **and** $(H=50, P_{cro}=0.9, P_{mut}=0.2)$

Group 1	Group 2
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	62743272
62743272	63346322

Part of the test result:

Method	Variance	df	t value	Pr > |t|
Summary	equal	18	-1.00	0.3306
Satterthwaite	unequal	9	-1.00	0.3434

Equal vaiance				
Method	**df(numerator)**	**df(denominator)**	**F value**	**Pr > F**
F	9	9	positive	<.0001

In the second table, the value of "Pr>F" is less than 0.05, which means the variance of the two groups is regarded equal. Then we turn to the first value of "Pr> |t|" in the first table, which is applicable for equal variance situation. It is larger than 0.05, which means that the difference between the two groups is not significant.

Test- 2: Average solutions under $(H=50, P_{cro}=0.7, P_{mut}=0.2)$ **and** $(H=50, P_{cro}=0.9, P_{mut}=0.2)$

Group 1	Group 2
69277453	69325195
68198929	69114353
68475077	69409038
69346057	70273566
69150848	70584365
69650159	67683498
69146095	69711526
69512226	69250028
69379201	67726477
69900782	70168125

Test result indicates that the difference between the two groups is not significant.

Method	Variance	df	t value	Pr > |t|
Summary	equal	18	-0.35	0.7338
Satterthwaite	unequal	13.609	-0.35	0.7351

Equal vaiance				
Method	**df(numerator)**	**df(denominator)**	**F value**	**Pr > F**
F	9	9	3.63	0.0683

Test- 3: Average solutions under $(H=80, P_{cro}=0.7, P_{mut}=0.2)$ **and** $(H=80, P_{cro}=0.9, P_{mut}=0.2)$

Group 1	Group 2
65815942	65668009
66181602	66110528
66246501	66915515
66175099	68348954
66031764	68037732
66211333	66535284
66081794	67638910
65867766	67323103
65969122	67288726
65987822	66652694

Test result indicates that the difference between the two groups is significant.

Method	Variance	df	t value	Pr > |t|
Summary	equal	18	-3.68	0.0017
Satterthwaite	unequal	9.5554	-3.68	0.0046

Equal vaiance				
Method	**df(numerator)**	**df(denominator)**	**F value**	**Pr > F**
F	9	9	32.38	<0.001

Appendix 3: The PRINCOMP Procedure

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	11.9762844	10.9073437	0.7485	0.7485
2	1.0689407	0.2246224	0.0668	0.8153
3	0.8443183	0.2309370	0.0528	0.8681
4	0.6133813	0.2591411	0.0383	0.9064
5	0.3542402	0.1583287	0.0221	0.9286
6	0.1959115	0.0548399	0.0122	0.9408
7	0.1410716	0.0077088	0.0088	0.9496
8	0.1333628	0.0134034	0.0083	0.9580
9	0.1199594	0.0196555	0.0075	0.9655
10	0.1003039	0.0030077	0.0063	0.9717
11	0.0972962	0.0093414	0.0061	0.9778
12	0.0879548	0.0060468	0.0055	0.9833
13	0.0819080	0.0122064	0.0051	0.9884
14	0.0697016	0.0051479	0.0044	0.9928
15	0.0645537	0.0137424	0.0040	0.9968
16	0.0508113		0.0032	1.0000

Appendix 4: Descriptive Statistics for Different Customer Groups

Index Name	Unit	Group1	Group2	Group3
Contribution from secondary industry	billion(RMB)	2240.86	2113.61	820.71
Contribution from tertiary industry	billion(RMB)	4146.41	2011.67	1385.62
Total export and import volume	billion(USD)	62.36	36.27	11.08
Foreign capital utilized	billion(USD)	8.05	4.26	0.96
Highway freight volume	billion× ton× kilometer	13.73	9.16	4.14
Railway freight volume	billion× ton× kilometer	23.09	18.05	10.03
Air freight volume by civil aviation	thousand × ton	1730.58	328.53	83.78
Industry output	billion(RMB)	1467.94	895.62	445.69

High-tech enterprises output	billion(RMB)	1389.49	458.91	89.96
Number of wholesalers and retailers	thousand	9.846	7.825	2.467
Sales volume of consumer goods	billion(RMB)	621.98	255.68	98.74
Disposable income of urban residents	RMB	26738	16789	8764
Urban population size	million	12.67	9.7	6.5
Business volume of postal delivery service and EDS	billion piece per year	80.56	75.95	67.86
Highway mileage	kilometer	20670	18906	10001

Reference

Abdinnour-Helm, S. (1998). A hybrid heuristic for the uncapacitated hub location problem. *European Journal of Operational Research*, 106(2), 489-499.

Abdinnour-Helm, S. & Venkataramanan, M. A. (1998). Solution approaches to hub location problems. *Annals of Operations Research*, 78, 31-50.

Agarwal, Y. K. (1989). An algorithm for designing survivable networks. *AT & T TECH. J.*, 68(3), 64-76.

Alba, E. (2005). *Parallel metaheuristics: a new class of algorithms.* Wiley-Interscience.

Albareda-Sambola, M., et al. (2005). A compact model and tight bounds for a combined location-routing problem. *Computers & operations research*, 32(3), 407-428.

Alminyana, M., et al. (1998). A new directed branching heuristic for the pq-median problem. *Location Science* 6(1), 1-23.

Altinkemer, K. & Yu, Z. (1992). Topological design of wide area communication networks. *Annals of Operations Research*, 36(1), 365-381.

Altinkemer, K. (1994). Topological design of ring networks. *Computers & operations research*, 21(4), 421-431.

Alumur, S. & Kara, B. Y. (2008a). Network hub location problems: The state of the art. *European Journal of Operational Research*, 190(1), 1-21.

Alumur, S. & Kara, B. Y. (2008b). A hub covering network design problem for cargo applications in Turkey. *Journal of the Operational Research Society*, 60(10), 1349-1359.

Alumur, S. A., et al. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B: Methodological*, 43(10), 936-951.

Anonymous (2008). Administrative Measures for Express Market. *Compile of New Registration and Law of P.R.China*, 9th Ed., China Legal Publishing House, Beijing.

Anonymous (2007). T*he First Statistic Survey on Express Delivery Service.* State Post Bureau of the People's Republic of China.

Armacost, A. P., et al. (2002). Composite Variable Formulations for Express Shipment Service Network Design. *Transportation Science*, 36(1), 1-20.

Armacost, A. P., et al. (2004). UPS optimizes its air network. *Interfaces*, 34(1), 15-25.

Aykin, T. (1994). Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research*, 79(3), 501-523.

Aykin, T. (1995a). Networking policies for hub-and-spoke systems with application to the air transportation system. *Transportation Science*, 29(3), 201-221.

Aykin, T (1995b). The hub location and routing problem. *European Journal of Operational Research*, 83(1), 200–219.

Bailey, E. E. (1991). *Deregulating the airlines*. The Massachusetts Institute of Technology, Fourth printing

Balakrishnan, A. & Graves, S. C. (1989). A composite algorithm for a concave-cost network flow problem. *Networks*, 19(2), 175-202.

Balakrishnan, A., et al. (1991). Models for planning capacity extension in local access telecommunication networks. *Annals of Operations Research*, 33(4), 237-284.

Balakrishnan, A. & Altinkemer, K. (1992). Using a hop-constrained model to generate alternative communication network design. *ORSA Journal on Computing*, 4(2), 192-205.

Balakrishnan, A., et al. (1994). A dual-based algorithm for multi-level network design. *Management Science, 40(5), 567-581.*

Balakrishnan, A., et al. (1995). A decomposition algorithm for local access telecommunications network extension planning. *Operations research*, 58-76.

Barnhart, C., et al. (2002). Network design for express shipment delivery. *Computational Optimization and Applications*, 21(3), 239-262.

Barreto, S., et al. (2007). Using clustering analysis in a capacitated location-routing problem. *European Journal of Operational Research*, 179(3), 968-977.

Bauer, P. W. (1987). Airline hubs: A study of determining factors and effects. *Economic Review-Federal Reserve Bank of Cleveland,4*, 13-19.

Bautista, J. & Pereira, J. (2007). A GRASP algorithm to solve the unicost set covering problem. *Computers & Operations Research*, 34(10), 3162-3173.

Baybars, I. & Edahl, R. H. (1988). A heuristic method for facility planning in telecommunications networks with multiple alternate routes. *Naval Research Logistics (NRL)*, 35(4), 503-528.

Beasley, J. E. & Nascimento, E. M. (1996). The vehicle routing-allocation problem: A unifying framework. *Top* ,4(1), 65-86.

Bednar, L. & Strohmeier, E. (1979). Lagerstandortoptimierung und Fuhrparkeinsatzplanung in der Konsumgüter-Industrie. *Mathematical Methods of Operations Research*, 23(4), 89-104.

Ben-Ayed, O. (2010). Hierarchical parcel distribution facility ground network design problem. *International Journal of Logistics Economics and Globalisation*, 2(3), 250-269.

Ben-Ayed, O. (2012). Parcel distribution network design problem. *Operational Research*, to be published, DOI10.1007/s12351-011-0118-2

Berman, O., et al. (2007). The transfer point location problem. *European journal of operational research*, 179(3), 978-989.

Berry, S. T. (1990). Airport presence as product differentiation. *The American Economic Review*, 80(2), 394-399.

Bhutta, K. S., et al. (2003). An integrated location, production, distribution and investment model for a multinational corporation. *International Journal of Production Economics*, 86(3), 201-216.

Blum, C. & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3), 268-308.

Blum, C. (2005). Beam-ACO hybridizing ant colony optimization with beam search: An application to open shop scheduling. *Computers & Operations Research*, 32(6), 1565-1591.

Blum, C. & Roli, A. (2008). Hybrid metaheuristics: an introduction. *Hybrid Metaheuristics*, 1-30.

Boland, N., et al. (2004). Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research*, 155(3), 638-653.

Bookbinder, J. H. & Reece, K. E. (1988). Vehicle routing considerations in distribution system design. *European Journal of Operational Research*, 37(2), 204-213.

Boorstyn, R. & Frank, H. (1977). Large-scale network topological optimization. *Communications, IEEE Transactions*, 25(1), 29-47.

Borenstein, S. (1989). Hubs and high fares: dominance and market power in the US airline industry. *The RAND Journal of Economics*, 344-365.

Borenstein, S. (1991). The dominant-firm advantage in multiproduct industries: Evidence from the US airlines. *The Quarterly Journal of Economics*, 106(4), 1237-1266.

Boventer, E. (1961). The relationship between transportation costs and location rent in transportation problems. *Journal of Regional Science*, 3(2), 27-40.

Bruns, A., et al. (2000). Restructuring of Swiss parcel delivery services. *OR Spectrum*, 22(2), 285-302.

Bryan, D. (1998). Extensions to the hub location problem: Formulations and numerical examples. *Geographical Analysis*, 30(4), 315-330.

Bryan, D. L. & O'Kelly, M. E. (1999). Hub-and-spoke Networks in Air Transportation: An Analytical Review. *Journal of Regional Science*, 39(2), 275-295.

Büdenbender, K., et al. (2000). A hybrid tabu search/branch-and-bound algorithm for the direct flight network design problem. *Transportation Science*, 34(4), 364-380.

Butler, R. V. & Huston, J. H. (1990). Airline service to non-hub airports ten years after deregulation. *Logistics and Transportation Review*, 26(1), 3-16.

Calik, H., et al. (2009). A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Computers & Operations Research*, 36(12), 3088-3096.

Campbell, J. F. (1993). Continuous and discrete demand hub location problems. *Transportation Research Part B: Methodological*, 27(6), 473-482.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2), 387-405.

Campbell, J. F. (1996). Hub location and the p-hub median problem. *Operations Research*, 923-935.

Campbell, J. F. (2002), Hub location problems. Chapt.12, pp. 373-402. In: Drezner Z, Hammacher H (eds) *Facility location: applications and theory*. Berlin, Springer

Campbell, J. F., et al. (2003). Solving hub arc location problems on a cluster of workstations. *Parallel Computing*, 29(5), 555-574.

Campbell, J. F., et al. (2005a). Hub arc location problems: part I: introduction and results. *Management Science*, 51(10), 1540-1555.

Campbell, J. F., et al. (2005b). Hub arc location problems: part II: formulations and optimal algorithms. *Management Science*, 51(10), 1556-1571.

Campbell, A. M., et al. (2007). The p-hub center allocation problem. *European journal of operational research*, 176(2), 819-835.

Campbell, J. F. (2009). Hub location for time definite transportation. *Computers & Operations Research*, 36(12), 3107-3116.

Canovas, L., et al. (2007). Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. *European Journal of Operational Research*, 179(3), 990-1007.

Cetiner, S., et al (2010). Hubbing and routing in postal delivery systems. *Annals of Operations Research*, 181(1), 109-124.

Chakravarthy, A., et al. (2009). Review of models and methodology for scheduling problems in USPS mail processing and distribution centers. *International Journal of Operational Research*, 5(4), 445-467.

Chang, S. G. & Gavish, B. (1993). Telecommunications network topological design and capacity extension: Formulations and algorithms. *Telecommunication Systems*, 1(1), 99-131.

Chao, I. (2002). A tabu search method for the truck and trailer routing problem. *Computers & Operations Research*, 29(1), 33-51.

Chen, J. F. (2007). A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega*, 35(2), 211-220.

Chen, H., et al. (2008). Network design for time constrained delivery. *Naval Research Logistics (NRL)*, 55(6), 493-515.

Chernev, A. & Kotler, P. (2008). *Strategic Marketing Management*. Brightstar Media, Incorporated.

Chestler, L. (1985). Overnight Air Express: Spatial Pattern, Competition and the Future in Small Package Delivery Service. *Transportation Quarterly*, 39(1), 59-71.

Chiarandini, M., et al. (2008). Very large-scale neighborhood search: Overview and case studies on coloring problems. *Hybrid Metaheuristics*, 117-150.

China Federation of Logistics & Purchasing (2008). *China Logistic Yearbook 2008*. China Material Press.

Chootinan, P. & Chen, A. (2006). Constraint handling in genetic algorithms using a gradient-based repair method. *Computers & operations research*, 33(8), 2263-2281.

Chou, Y. H. (1990). The hierarchical-hub model for airline networks. *Transportation Planning and Technology*, 14(4), 243-258.

Colome, R., et al. (2003). A new chance-constrained maximum capture location problem. *Annals of Operations Research* ,122(1), 121-139.

Cominetti, R. & Ortega, F. (1997). *A branch & bound method for minimum concave cost network flows based on sensitivity analysis*. Universidad de chile.

Committee for China Transportation Yearbook (2011). *Year Book of China Transportation & Communication*. Year Book House of China Transporation & Communication, Beijing.

Contreras, I., et al. (2011). Branch and price for large-scale capacitated hub location problems with single assignment. *INFORMS Journal on Computing*, 23(1), 41-55.

Cooper, L. (1963). Location-allocation problems. *Operations Research*, 11(3), 331-343.

Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM review*, 6(1), 37-53.

Costa, A. M., et al. (2011). Two-level network design with intermediate facilities: An application to electrical distribution systems. *Omega*, 39(1), 3-13.

Crainic, T. G. & Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 97(3), 409-438.

Crainic, T. G. (2000). Service network design in freight transportation. *European Journal of Operational Research*, 122(2), 272-288.

Crainic, T. (2003). Long-haul freight transportation. *Handbook of transportation science*, 451-516.

Crainic, T. G. & Kim, K. H. (2007). Intermodal transportation. *Transportation*, 14, 467-537.

Croxton, K. L., et al. (2003). A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems. *Management Science*, 49(9), 1268-1273.

Croxton, K. L., et al. (2007). Variable disaggregation in network flow problems with piecewise linear costs. *Operations research*, 55(1), 146-157.

Cunha, C. B. & Silva, M. R. (2007). A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil. *European Journal of Operational Research*, 179(3), 747-758.

Current, J., et al. (1990). Multiobjective analysis of facility location decisions. *European Journal of Operational Research*, 49(3), 295-307.

Current, J. R. & Schilling, D. A. (1994). The median tour and maximal covering tour problems: Formulations and heuristics. *European Journal of Operational Research*, 73(1), 114-126.

da Graca Costa, M., et al. (2008). Capacitated single allocation hub location problem- A bi-criteria approach. *Computers & Operations Research*, 35(11), 3671-3695.

Dasci, A. & Laporte, G. (2004). An analytical approach to the facility location and capacity acquisition problem under demand uncertainty. *Journal of the Operational Research Society*, 56(4), 397-405.

Daskin, M. S. & Panayotopoulos, N. D. (1989). A Lagrangian relaxation approach to assigning aircraft to routes in hub and spoke networks. *Transportation Science*, 23(2), 91-99.

Daskin, M. S. (1995). *Network and discrete location: models, algorithms, and applications.* Wiley-Interscience series in discrete mathematics and optimization, Wiley, New York,

Dawkins, R. (2006). *The selfish gene.* Oxford University Press, USA.

de Assis Correa, F., et al. (2007). Hybrid heuristics for the probabilistic maximal covering location-allocation problem. *Operational Research*, 7(3), 323-343.

De Camargo, R. S., et al. (2008). Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers & Operations Research*, 35(4), 1047-1064.

De Jong, K. & Spears, W. (1991). An analysis of the interacting roles of population size and crossover in genetic algorithms. *Parallel problem solving from nature*, 496, 38-47.

De Jong, K. A. (2012). *Evolutionary computation: a unified approach.* Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion, ACM, 737-750.

Della Croce, F. & Tkindt, V. (2002). A recovering beam search algorithm for the one-machine dynamic total completion time scheduling problem. *Journal of the Operational Research Society*, 53(11), 1275-1280.

Delorme, X., et al. (2004). GRASP for set packing problems. *European Journal of Operational Research*, 153(3), 564-580.

Dibb, S. (1999). Criteria guiding segmentation implementation: reviewing the evidence. *Journal of Strategic Marketing*, 7(2), 107-129.

Dobson, G. & Lederer, P. J. (1993). Airline scheduling and routing in a hub-and-spoke system. *Transportation Science*, 27(3), 281-297.

Domschke, W.&Krispin, G. (1997). Location and layout planning. *OR Spectrum*, 19(3), 181-194.

Domschke, W. & Drexl, A. (1998). *Einführung in Operations-Research.* Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL)

Domschke, W. & Krispin, G. (1999). Zur wirtschaftlichen Effizienz von Hub-and-Spoke-Netzen. *Logistikforschung*, 279-304.

Domschke, W. (2010). *Logistik: Rundreisen und Toure*n. 2 Oldenbourg Verlag.

Drezner, Z. (1995). On the conditional p-median problem. *Computers & operations research*, 22(5), 525-530.

Drezner, Z. & Hamacher, H. W. (2001). *Facility location: applications and theory*. Springer.

Drezner, T., et al. (2002). Solving the multiple competitive facilities location problem. *European Journal of Operational Research*, 142(1), 138-151.

Ebery, J., et al. (2000). The capacitated multiple allocation hub location problem: Formulations and algorithms. *European Journal of Operational Research*, 120(3), 614-631.

Ebery, J. (2001). Solving large single allocation p-hub problems with two or three hubs. *European Journal of Operational Research*, 128(2), 447-458.

Eiselt, H. A. (2007). Locating landfills- optimization vs. reality. *European journal of operational research*, 179(3), 1040-1049.

Ernst, A. T. & Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location science*, 4(3), 139-154.

Ernst, A. T. & Krishnamoorthy, M. (1998a). Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research*, 104(1), 100-112.

Ernst, A. T. & Krishnamoorthy, M. (1998b). An exact solution approach based on shortest-paths for p-hub median problems. *INFORMS Journal on Computing*, 10(2), 149-162.

Ernst, A. T. & Krishnamoorthy, M. (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, 86, 141-159.

Ernst, A. T., et al. (2005). Reformulations and computational results for uncapacitated single and multiple allocation hub covering problems. *Unpublished Report, CSIRO Mathematical and Information Sciences, Australia.*

Ernst, A. T., et al. (2009). Uncapacitated single and multiple allocation p-hub center problems. *Computers & Operations Research*, 36(7), 2230-2241.

Everitt B. S., et al. (2011).*Cluster analysis.* 5th Ed.,Wiley

Farahani, R. Z., et al. Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling* , 34(7), 1689-1709.

Feo, T. A. & Resende, M. G. C. (1989). A probabilistic heuristic for a computationally difficult set covering problem. *Operations research letters*, 8(2), 67-71.

Festa, P. & Resende, M. G. C. (2009). An annotated bibliography of GRASP- Part I: Algorithms. *International Transactions in Operational Research*, 16(1), 1-14.

Festa, P. & Resende, M. G. C. (2009). An annotated bibliography of GRASP- Part II: Application. *International Transactions in Operational Research*, 16(2), 131-172.

Fleurent, C. & Ferland, J. A. (1994). Genetic hybrids for the quadratic assignment problem. *Quadratic assignment and related problems*, 16, 173-187.

Flynn, J. & Ratick, S. (1988). A multiobjective hierarchical covering model for the essential air services program. *Transportation Science*, 22(2), 139-147.

Focacci, F., et al. (2003). Local search and constraint programming. *Handbook of metaheuristics*, 369-403.

Galinier, P. & Hao, J. K. (1999). Hybrid evolutionary algorithms for graph coloring. *Journal of combinatorial optimization*, 3(4), 379-397.

Garfinkel, R. S., et al. (1996). Optimal use of hub facilities: A two-hub model with fixed arc costs. *Top*, 4(2), 331-343.

Gavish, B. & Altinkemer, K. (1990). Backbone network design tools with economic tradeoffs. *ORSA Journal on Computing*, 2(3), 236-252.

Gavish, B. (1991). Topological design of telecommunication networks-local access design methods. *Annals of Operations Research*, 33(1), 17-71.

Gavish, B. (1992). Topological design of computer communication networks- the overall design problem. *European Journal of Operational Research*, 58(2), 149-172.

Gavish, B. & Suh, M. W. (1992). Configuration of fully replicated distributed database system over wide area networks. *Annals of Operations Research*, 36(1), 167-191.

Gelareh, S., et al. (2010). Liner shipping hub network design in a competitive environment. *Transportation Research Part E: Logistics and Transportation Review*, 46(6), 991-1004.

Gen, M. & Cheng, R. (2000). *Genetic algorithms and engineering optimization.* Wiley-interscience.

Gendreau, M., et al. (1997). The covering tour problem. *Operations Research*, 45(4), 568-576.

Gendreau, M. & Potvin, J. Y. (2010). *Handbook of metaheuristics.* Springer.

Ghosh, D. (2003). Neighborhood search heuristics for the uncapacitated facility location problem. *European Journal of Operational Research*, 150(1), 150-162.

Gilbert A. Churchill, J. D. I. (2009). *Marketing Research: Methodological Foundations.* 10th Ed. South-Western Cengage Learning.

Glover, F. & Kochenberger, G. A. (2003). *Handbook of metaheuristics.* Springer.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley.

Goldberg, D. E. & Voessner, S. (1999). Optimizing global/local search hybrids. *Urbana*, 51, 220-228.

Gong, D., et al. (1997). Hybrid evolutionary method for capacitated location-allocation problem. *Computers & industrial engineering*, 33(3-4), 577-580.

Grefenstette, J. J. (1986). Optimization of control parameters for genetic algorithms. *Systems, Man and Cybernetics, IEEE Transactions*, 16(1), 122-128.

Groothedde, B. & Tavasszy, L. A. (1999). Optimalisatie van terminal locaties in een multimodaal netwerk met simulated annealing. *Proceedings van de Vervoerslogistieke Werkdagen*, 43-57.

Groothedde, B., et al. (2005). Towards collaborative, intermodal hub networks:: A case study in the fast moving consumer goods market. *Transportation Research Part E: Logistics and Transportation Review*, 41(6), 567-583.

Grove, P. G. & O'Kelly, M. E. (1986). Hub networks and simulated schedule delay. *Regional Science*, 59(1), 103-119.

Grünert, T. & Sebastian. H.-J. (2000). Planning models for long-haul operations of postal and express shipment companies. *European Journal of Operational Research*, 122(2), 289-309.

Guisewite, G. M. & Pardalos, P. M. (1990). Minimum concave-cost network flow problems: Applications, complexity, and algorithms. *Annals of Operations Research*, 25(1), 75-99.

Gunnarsson, H., et al. (2005). A combined terminal location and ship routing problem. *Journal of the Operational Research Society*, 57(8), 928-938.

Hale, T. S. & Moberg, C. R. (2003). Location science research: a review. *Annals of Operations Research*, 123(1), 21-35.

Hall, R. W. (1989). Configuration of an overnight package air network. *Transportation Research Part A: General* ,23(2), 139-149.

Hall, R. W. (1996). Pickup and delivery systems for overnight carriers. *Transportation Research Part A: Policy and Practice*, 30(3), 173-187.

Hall, R. W. (2001). Truck scheduling for ground to air connectivity. *Journal of Air Transport Management*, 7(6), 331-338.

Hamacher, H. W., et al. (2004). Adapting polyhedral properties from facility to hub location problems. *Discrete Applied Mathematics*, 145(1), 104-116.

Hamacher, H. W. & Meyer, T. (2006). Hub cover and hub center problems. *Departament of Mathematics, University of Kaiserslautern, Germany*. 1-18.

Hansen, P. & Mladenovic, N. (1997). Variable neighborhood search for the p-median. *Location Science*, 5(4), 207-226.

Hassanzadeh, A., et al. (2009). Location-Routing Problem. *Facility Location*, 395-417.

Haupt, R. L., et al. (2004). *Practical genetic algorithms*. Wiley Online Library.

He, L. & Mort, N. (2000). Hybrid genetic algorithms for telecommunications network back-up routing. *BT Technology Journal*, 18(4), 42-50.

Hekmatfar, M. & Pishvaee, M. (2009). Hub Location Problem. *In* "Facility Location: Concepts, Models, Algorithms and Case Studies" (R. Z. Farahani & M. Hekmatfar, Eds.), pp. 243-270. Physica-Verlag, Heidelburg.

Hertz, A. & Widmer, M. (2003). Guidelines for the use of meta-heuristics in combinatorial optimization. *European Journal of Operational Research*, 151(2), 247-252.

Hertz, A. & Kobler, D. (2000). A framework for the description of evolutionary algorithms. *European Journal of Operational Research*, 126(1), 1-12.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* University of Michigan press.

Holmberg, K. (1994). Solving the staircase cost facility location problem with decomposition and piecewise linearization. *European Journal of Operational Research*, 75(1), 41-61.

Holmberg, K. & Ling, J. (1997). A Lagrangean heuristic for the facility location problem with staircase costs. *European Journal of Operational Research*, 97(1), 63-74.

Holmqvist, K., et al. (1998). A GRASP algorithm for the single source uncapacitated minimum concave-cost network flow problem. *Network design: Connectivity and facilities location*, 40, 131-142.

Horner, M. W. & O'Kelly, M. E. (2001). Embedding economies of scale concepts for hub network design. *Journal of Transport Geography*, 9(4), 255-265.

Horst, R. & Hoang, T. (1996). *Global optimization: Deterministic approaches.* Springer Verlag.

Ibaraki, T. & Nakamura, K. (2006). Packing problems with soft rectangles. *Hybrid Metaheuristics*, 13-27.

Ishfaq, R. & Sox, C. R. Hub location-allocation in intermodal logistic networks. *European Journal of Operational Research*, 210(2), 213-230.

Jacobsen, S. K. & Madsen, O. B. G. (1980). A comparative study of heuristics for a two-level routing-location problem. *European Journal of Operational Research*, 5(6), 378-387.

Jägersküpper, J. & Storch, T. (2007). When the plus strategy outperforms the comma strategy-and when not. *Proc. of FOCI* 7.

Jaillet, P., et al. (1996). Airline network design and hub location problems. *Location Science*, 4(3), 195-212.

Jansen, B., et al. (2004). Operational planning of a large-scale multi-modal transportation system. *European Journal of Operational Research*, 156(1), 41-53.

Jansen, T., et al. (2005). On the choice of the offspring population size in evolutionary algorithms. *Evolutionary Computation*, 13(4), 413-440.

Jaszkiewicz, A. (2002). Genetic local search for multi-objective combinatorial optimization. *European journal of operational research*, 137(1), 50-71.

Jourdan, L., et al. (2009). Hybridizing exact methods and metaheuristics: A taxonomy. *European Journal of Operational Research*, 199(3), 620-629.

Jourquin, B., et al. (1999). Freight bundling network models: Methodology and application. *Transportation Planning and Technology*,23(2), 157-177.

Jozefowiez, N., et al. (2009). An evolutionary algorithm for the vehicle routing problem with route balancing. *European Journal of Operational Research*, 195(3), 761-769.

Kalfakakou, R. & Tsouros, C. C. (2001). Determining the size and location of suppliers for a minimum total distribution route problem. *International Journal of Production Economics*, 71(1), 401-405.

Kamakura, M. W. W. (1999). *Market segmentation: conceptual and methodological foundations.* 2nd. Ed.. Kluwer Academic Publishers, Massachusetts.

Kamoun, M. & Hall, R. W. (1996). Design of express mail services for metropolitan regions. *Journal of Business Logistics*, 17(2), 265-302.

Kara, B. Y. (1999). *Modeling and analysis of issues in hub location problems.* Doctor of Philosophy Thesis, Bilkent University, Ankara, Turkey.

Kara, B. Y. & Tansel, B. C. (2000). On the single-assignment p-hub center problem. *European Journal of Operational Research*, 125(3), 648-655.

Kara, B. Y. & Tansel, B. C. (2001). The latest arrival hub location problem. *Management Science*, 47(10), 1408-1420.

Kara, B. Y. & Tansel, B. C. (2003). The single-assignment hub covering problem: Models and linearizations. *Journal of the Operational Research Society*, 54, 59-64.

Kim, D. & Barnhart, C. (1997). Multimodal express shipment service design: Models and algorithms. *Computers & industrial engineering*, 33(3-4), 685-688.

Kim, D., et al. (1999). Multimodal express package delivery: A service network design application. *Transportation Science*, 33(4), 391-407.

Kim, D. & Pardalos, P. M. (2000). Dynamic slope scaling and trust interval techniques for solving concave piecewise linear network flow problems. *Networks*, 35(3), 216-222.

Kim, D. & Pardalos, P. M. (2000). A Dynamic Domain Contraction Algorithm for Nonconvex Piecewise Linear Network Flow Problems. *Journal of Global Optimization*, 17(1), 225-234.

Kim, T. et al., (2007). Intermodal Transportation. *In "Handbooks in Operations Research and Management Science* " (C. B. G. Laporte, Ed.), Vol. 14, pp. 467-529. Elsevier, Amsterdam.

Kimms, A. (2006). Economies of scale in hub & spoke network design models: We have it all wrong. *Perspectives on Operations Research Essays in Honor of Klaus Neumann, Wiesbaden, Gabler, S*, 293-317.

Kleinrock, L. & Kamoun, F. (1980). Optimal clustering structures for hierarchical topological design of large computer networks. *Networks*, 10(3), 221-248.

Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research*, 53(1), 25-37.

Klincewicz, J.G. (1992) Avoiding local optima in the P-hub location problem using tabu search and grasp. *Annals of Operations Research*, 40 (1), 283–302.

Klincewicz, J. G. (1996). A dual algorithm for the uncapacitated hub location problem. *Location Science*, 4(3), 173-184.

Klincewicz, J. G. (1998). Hub location in backbone/tributary network design: a review. *Location Science*, 6(1-4), 307-335.

Klincewicz, J. G. (2002). Enumeration and search procedures for a hub location problem with economies of scale. *Annals of Operations Research*, 110(1), 107-122.

Klose, A. & Drexl, A. (2005). Facility location models for distribution system design. *European Journal of Operational Research*, 162(1), 4-29.

Kotler P. et al., (2009). *Marketing Management.* 13th Ed. Prentice Hall, Edinburgh.

Kratica, J., et al. (2001). Solving the simple plant location problem by genetic algorithm. *RAIRO-Operations Research*, 35(1), 127-142.

Kratica, J.&Stanimirovic, Z. (2006). Solving the uncapacitated multiple allocation p-hub center problem by genetic algorithm. *Asia-Pacific Journal of Operational Research*, 23(04), 425-437.

Kratica, J., et al. (2007). Two genetic algorithms for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, 182(1), 15-28.

Kreinovich, V., et al. (1993). Genetic algorithms: What fitness scaling is optimal? *Cybernetics and Systems*, 24(1), 9-26.

Kreutzberger, E. D. (2008). Distance and time in intermodal goods transport networks in Europe: A generic approach. *Transportation Research Part A: Policy and Practice*, 42(7), 973-993.

Kuby, M. J. & Gray, R. G. (1993). The hub network design problem with stopovers and feeders: The case of Federal Express. *Transportation Research Part A: Policy and Practice*, 27(1), 1-12.

Labbe, M. & Yaman, H. (2004). Projecting the flow variables for hub location problems. *Networks*, 44(2), 84-93.

Labbe, M., et al. (2005). A branch and cut algorithm for hub location problems with single assignment. *Mathematical programming*, 102(2), 371-405.

Labbe, M., et al. (2005). Locating median cycles in networks. *European Journal of Operational Research*, 160(2), 457-470.

Labbe, M.&Yaman, H. (2008). Solving the hub location problem in a star-star network. *Networks*, 51(1), 19-33.

Lamar, B. W. (1993). An improved branch and bound algorithm for minimum concave cost network flow problems. *Journal of Global Optimization*, 3(3), 261-287.

Laporte, G. & Nobert, Y. (1981). An exact algorithm for minimizing routing and operating costs in depot location. *European Journal of Operational Research*, 6(2), 224-226.

Larsson, T., et al. (1994). A Lagrangean heuristic for the capacitated concave minimum cost network flow problem. *European Journal of Operational Research*, 78(1), 116-129.

Lawrence, R. M. & Pengilly, P. J. (1969). The number and location of depots required for handling products for distribution to retail stores in South-East England. *OR*, 23-32.

Lee, C. Y. (1993). An algorithm for the design of multitype concentrator networks. *Journal of the Operational Research Society*, 44 (5), 471-482.

Lee, Y., et al. (1996). A hub location problem in designing digital data service networks: Lagrangian relaxation approach. *Location Science*, 4(3), 185-194.

Leung, J. M. Y., et al. (1990). Routing in point-to-point delivery systems: formulations and solution heuristics. *Transportation science*, 24(4), 245-260.

Lim, M. H., et al. (2000). Efficient genetic algorithms using simple genes exchange local search policy for the quadratic assignment problem. *Computational Optimization and Applications*, 15(3), 249-268.

Lin, C. K. Y., et al. (2002). A location-routing-loading problem for bill delivery services. *Computers & industrial engineering*, 43(1), 5-25.

Lin, C. C., et al. (2003). The economic effects of center-to-center directs on hub-and-spoke networks for air express common carriers. *Journal of Air Transport Management*, 9(4), 255-265.

Lin, C. C. & Chen, S. H. (2004). The hierarchical network design problem for time-definite express common carriers. *Transportation Research Part B: Methodological*, 38(3), 271-283.

Lin, C. K. Y. & Kwok, R. C. W. (2006). Multi-objective metaheuristics for a location-routing problem with multiple use of vehicles on real data and simulated data. *European journal of operational research*, 175(3), 1833-1849.

Lin, C. C. & Chen, S. H. (2008). An integral constrained generalized hub-and-spoke network design problem. *Transportation Research Part E: Logistics and Transportation Review*, 44(6), 986-1003.

Lin, C. C. (2010).The integrated secondary route network design model in the hierarchical hub-and-spoke network for dual express services. *International Journal of Production Economics*, 123(1), 20-30.

Lin, C. C. & Lee, S. C. (2010) The competition game on hub network design. *Transportation Research Part B: Methodological*, 44(4), 618-629.

Liu, B. (2009). Facility Location Problem. *Theory and Practice of Uncertain Programming*, 157-165.

Ljubic, I. (2007). A hybrid VNS for connected facility location. *Hybrid Metaheuristics*, 4771, 157-169.

Lowe T.J & Sim T. The hub covering flow problem. Working paper. It will be appearing in the Journal of the Operational Research Society.

Macharis, C. & Bontekoning, Y. M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, 153(2), 400-416.

Madsen, O. B. G. (1983). Methods for solving combined two level location-routing problems of realistic dimensions. *European Journal of Operational Research*, 12(3), 295-301.

Maniezzo, V. (1999). Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem. *INFORMS Journal on Computing*, 11(4), 358-369.

Maranzana, F. E. (1964). On the location of supply points to minimize transport costs. *OR*, 261-270.

Marianov, V. & Serra, D. (2003). Location models for airline hubs behaving as M/D/c queues. *Computers & Operations Research*, 30(7), 983-1003.

Marin, A. (2005a). Formulating and solving splittable capacitated multiple allocation hub location problems. *Computers & operations research*, 32(12), 3093-3109.

Marin, A. (2005b). Uncapacitated Euclidean hub location: Strengthened formulation, new facets and a relax-and-cut algorithm. *Journal of Global Optimization*, 33(3), 393-422.

Marin, A., et al. (2006). New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research*, 172(1), 274-292.

Macharis, C. & Bontekoning, Y. M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, 153(2), 400-416.

Marti, R. (2003). Multi-start methods. *Handbook of metaheuristics*, Chapter 12, 355-368.

Martin, J. C. & Roman, C. (2004). Analyzing competition for hub location in intercontinental aviation markets. *Transportation Research Part E: Logistics and Transportation Review*, 40(2), 135-150.

Mayer, G. (2001). *Strategische Logistikplanung von Hub-&-Spoke-Systemen*. Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL)

Mayer, G. & Wagner, B. (2002). HubLocator: an exact solution method for the multiple allocation hub location problem. *Computers & Operations Research*, 29(6), 715-739.

McDonald, M. & Dunbar, I. (2004). *Market Segmentation: How to do it, how to profit from it.* Elserier Butterworth-Heinemann.

McGinnis, M. A. (1989). A comparative evaluation of freight transportation choice models. *Transportation Journal* , 29(2), 36-46.

McShan, W. S. (1986). *An economic analysis of the hub-and-spoke routing strategy in the airline industry.* Thesis (Ph.D.), Northwestern University, 1986.

Melachrinoudis, E. & Min, H. (2000). The dynamic relocation and phase-out of a hybrid, two-echelon plant/warehousing facility: A multiple objective approach. *European Journal of Operational Research*, 123(1), 1-15.

Melechovsky, J., et al. (2005). A metaheuristic to solve a location-routing problem with non-linear costs. *Journal of Heuristics*, 11(5), 375-391.

Merz, P. & Freisleben, B. (1997). Genetic local search for the TSP: new results. Conference Publications, *Evolutionary Computation*, 1997, IEEE International Conference, 159-164.

Merz, P. (2000). Fitness landscape analysis and memetic algorithms for the quadratic assignment problem. *Evolutionary Computation, IEEE Transactions*, 4(4), 337-352

Meuffels, W. J. M., et al. (2010).Enriching the tactical network design of express service carriers with fleet scheduling characteristics. *Flexible Services and Manufacturing Journal*, *22*(1), 3-35.

Michalewicz, Z. & Schoenauer, M. (1996). Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary computation*, 4(1), 1-32.

Min, H., et al. (1998). Combined location-routing problems: A synthesis and future research directions. *European Journal of Operational Research*, 108(1), 1-15.

Minoux, M. (1989). Networks synthesis and optimum network design problems: Models, solution methods and applications. *Networks*, 19(3), 313-360.

Mobasheri, F., et al. (1989). Scenario planning at southern California Edison. *Interfaces*, 19(5), 31-44.

Monma, C. & Sheng, D. (1986). Backbone network design and performance analysis: A methodology for packet switching networks. Selected Areas in Communications, IEEE Journal, 4(6), 946-965.

Morris, G. M., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14), 1639-1662.

Morrison, S. & Winston, C. (1986). *The economic effects of airline deregulation*. Brookings Institution Press.

Moscato, P. (1989). On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. *Caltech concurrent computation program, C3P Report* ,826, 1989.

Mourgaya, M & Vanderbeck, F. (2007). Column generation based heuristic for tactical planning in multi-period vehicle routing. *European Journal of Operational Research*, 183(3), 1028-1041.

Mulvey, J. M. (1996). Generating scenarios for the Towers Perrin investment system. *Interfaces*, 26(2), 1-15.

Murata,T. & Ishibuchi, H. (1994) Performance evaluation of genetic algorithms for flow shop scheduling problems, *Proceedings of the First IEEE Conference on Evolutionary Computation*, 2, 812-817.

Murty, K. G. & Djang, P. A. (1999). The US Army national guard's mobile training simulators location and routing problem. *Operations Research*, 47(2), 175-182.

Myers, J. H. & Mullet G. M. (2003). Managerial Applications of Multivariate Analysis in Marketing. *American Marketing Association*.

Nagy, G. & Salhi, S. (1996a). Nested heuristic methods for the location-routing problem. *Journal of the Operational Research Society* 47, 1166–1174.

Nagy, G. & Salhi, S. (1996b). A nested location-routing heuristic using route length estimation. *Studies in Locational Analysis* 10, 109–127.

Nagy, G. & Salhi, S. (1998). The many-to-many location-routing problem. *Top*, 6(2), 261-275.

Nagy, G. & Salhi, S. (2007). Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2), 649-672.

Nahapetyan, A. & Pardalos, P. M. (2007). A bilinear relaxation based algorithm for concave piecewise linear network flow problems. *Journal of Industrial and Management Optimization*, 3(1), 71.

Nambiar, J. M., et al. (1981). A large scale location-allocation problem in the natural rubber industry. *European Journal of Operational Research*, 6(2), 183-189.

Nambiar, J. M., et al. (1989). Plant location and vehicle routing in the Malaysian rubber smallholder sector: A case study. *European journal of operational research*, 38(1), 14-26.

Newman, P. (1988). A fast packet switch for the integrated services backbone network. *Selected Areas in Communications, IEEE Journal* , 6(9), 1468-1479.

Noviello, K., et al. (1996). A comparison of the air passenger and air cargo industries with respect to hub locations. *Great Lakes Geographer*, 3, 75-85.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, 32(3), 393-404.

O'Kelly, M. E. & Lao, Y. (1991). Mode Choice in a Hub-and-spoke Network: A Zero-one Linear Programming Approach. *Geographical Analysis*, 23(4), 283-297.

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

O'Kelly, M. E. & Miller, H. J. (1994). The hub network design problem: a review and synthesis. *Journal of Transport Geography*, 2(1), 31-40.

O'Kelly, M. E. & Skorin-Kapov, J. (1996). Hub network design with single and multiple allocation: A computational study. *Location Science*, 4(3), 125-138.

O'Kelly, M. E. (1998). A geographer's analysis of hub-and-spoke networks. *Journal of Transport Geography*, 6(3), 171-186.

O'Kelly, M. E. & Bryan, D. L. (1998). Hub location with flow economies of scale. *Transportation Research Part B: Methodological*, 32(8), 605-616.

O'kelly, M. E. & Bryan, D. (2002). Interfacility interaction in models of hub and spoke networks. *Journal of Regional Science*, 42(1), 145-164.

Osman, I. H. & Laporte, G. (1996). Metaheuristics: A bibliography. *Annals of Operations Research*, 63(5), 511-623.

Osman, M. S., et al. (2004). A solution to the optimal power flow using genetic algorithm. *Applied mathematics and computation*, 155(2), 391-405.

Owen, S. H.&Daskin, M. S. (1998). Strategic facility location: A review. *European Journal of Operational Research*, 111(3), 423-447.

Oxford Economics (2009). The Impact of the Express Delivery Industry on the Global Economy. URL: https://canadiancourier.org/uploads/Impact_of_Express_Industry_-_Global_Economy_-2009.pdf (Access on Nov. 4 2012)

Pamuk, F. S. & Sepil, C. (2001). A solution to the hub center problem via a single-relocation algorithm with tabu search. *IIE Transactions*, 33(5), 399-411.

Perl, J. & Daskin, M. S. (1985). A warehouse location-routing problem. *Transportation Research Part B: Methodological* , 19(5), 381-396.

Pfohl, H. C. (2004). *Logistikmanagement: konzeption und funktionen*. Ed.2, Berlin et al.

Pirkul, H., et al. (1991). The hierarchical network design problem: a new formulation and solution procedures. *Transportation science*, 25(3), 175-182.

Pirkul, H. & Nagarajan, V. (1992). Locating concentrators in centralized computer networks. *Annals of Operations Research*, 36(1), 247-261.

Pirkul, H. & Schilling, D. A. (1998). An efficient procedure for designing single allocation hub and spoke systems. *Management Science*, 44, 235-242.

Podnar, H., et al. (2002). Network cost minimization using threshold-based discounting. *European Journal of Operational Research*, 137(2), 371-386.

Podnar, H. & Skorin-Kapov, J. (2003). Genetic algorithm for network cost minimization using threshold based discounting. *Journal of Applied Mathematics & Decision Sciences*, 7(4), 207-228.

Porter, M. E. (1998). *Competitive Advantage, Creating and Sustaining Superior Performance*. The Free Press, New York.

Powell, W. B. (1986). A local improvement heuristic for the design of less-than-truckload motor carrier networks. *Transportation Science*, 20(4), 246-257.

Puchinger, J. & Raidl, G. (2005). Combining metaheuristics and exact algorithms in combinatorial optimization: A survey and classification. *Artificial intelligence and knowledge engineering applications: a bioinspired approach*, 113-124.

Qu, B. & Weng, K. (2009). Path relinking approach for multiple allocation hub maximal covering problem. *Computers & Mathematics with Applications*, 57(11-12), 1890-1894.

Racunica, I. & Wynter, L. (2005). Optimal location of intermodal freight hubs. *Transportation Research Part B: Methodological* , 39(5), 453-477.

Raidl, G. (2006). A unified view on hybrid metaheuristics. *Hybrid Metaheuristics*, 1-12.

Raidl, G. & Puchinger, J. (2008). Combining (integer) linear programming techniques and metaheuristics for combinatorial optimization. *Hybrid Metaheuristics*, 114, 31-62.

Rechenberg, I. (1971). *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Dissertation (PhD Thesis), Technical University of Berlin, Department of Process Engineering

Repolho, H. M., et al. (2010). Optimum location of motorway interchanges: Users-perspective. *Journal of Transportation Engineering*, 136, 956.

Rearch Panel from Department of Commerce (2009). *Research on Development of Express Market in China*. China Machine Press, Beijing.

Resende, M. & Ribeiro, C. (2003). Greedy randomized adaptive search procedures. *Handbook of metaheuristics*, 219-249.

Resende, M. G. C.&Werneck, R. F. (2006). A hybrid multistart heuristic for the uncapacitated facility location problem. *European Journal of Operational Research*, 174(1), 54-68.

Resende, M. G. C. & Werneck, R. F. (2007). A fast swap-based local search procedure for location problems. *Annals of Operations Research*, 150(1), 205-230.

ReVelle, C. (1997). A perspective on location science. *Computers & Operations Research*, 5(1), 3 -13.

ReVelle, C. S. & Eiselt, H. A. (2005). Location analysis: A synthesis and survey. *European Journal of Operational Research*, 165(1), 1-19.

ReVelle, C., et al. (2007). Location models for ceding market share and shrinking services. *Omega 35(5)*, 533-540.

Revelle, C. S., et al. (2008). A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184(3), 817-848.

Reynolds-Feighan, A. (2001). Traffic distribution in low-cost and full-service carrier networks in the US air transportation market. *Journal of Air Transport Management*, 7(5), 265-275.

Rodriguez, V., et al. (2007). Hub location under capacity constraints. *Transportation Research Part E: Logistics and Transportation Review*, 43(5), 495-505.

Rodriguez-Martin, I. & Salazar-Gonzalez, J. J. (2008). Solving a capacitated hub location problem. *European Journal of Operational Research*, 184(2), 468-479.

Romesburg, H.C (2004), *Cluster analysis for researchers*, Lulu press

Roy, J. & Crainic, T. G. (1992). Improving intercity freight routing with a tactical planning model. *Interfaces*, 22(3), 31-44.

Runarsson, T. P. & Yao, X. (2000). Stochastic ranking for constrained evolutionary optimization. *Evolutionary Computation, IEEE Transactions on*, 4(3), 284-294.

Saha, D. & Mukherjee, A. (1995). Design of hierarchical communication networks under node/link failure constraints. *Computer Communications*, 18(5), 378-383.

Sahni, S. & Gonzalez, T. (1976). P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3), 555-565.

Salhi, S. & Rand, G. K. (1989). The effect of ignoring routes when locating depots. *European Journal of Operational Research*, 39(2), 150-156.

Salhi, S. & Nagy, G. (1999). Consistency and robustness in location-routing. *Studies in Locational Analysis* (13), 3-19.

Salhi, S. & Gamal, M. D. H. (2003). A Genetic algorithm based approach for the uncapacitated continuous location-allocation problem. *Annals of Operations Research*, 123(1), 203-222.

Salhi, S. & Nagy, G. (2009). Local improvement in planar facility location using vehicle routing. *Annals of Operations Research*, 167(1), 287-296.

Santoso, T., et al. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1), 96-115.

Sasaki, M., et al. (1999). On the selection of hub airports for an airline hub-and-spoke system. *Computers & operations research* , 26(14), 1411-1422.

Sasaki, M. & Fukushima, M. (2003). On the hub-and-spoke model with arc capacity constraints. *Journal of the Operations Research Society of Japan-Keiei Kagaku*, 46(4), 409-428.

Schaffer, J. D., et al. (1989). Conference paper appear in Combinations of Genetic Algorithms and Neural Networks, 1992., COGANN-92. *International Workshop*, 1-37.

Serra, D.& ReVelle, C. (1993). The pq-median problem, location and districting of hierarchical facilities.*Location Science1*, (4), 299-312.

Serra, D. & ReVelle, C (1994). The pq-median problem: location and districting of hierarchi cal facilities.-II. Heuristic solution methods. *Location Science*, 2 (2), 63-82.

Silk, A. J. (2006). *What is marketing?* Harvard Business Press.

Silva, M. R. & Cunha, C. B. (2009). New simple and efficient heuristics for the uncapacitated single allocation hub location problem. *Computers & Operations Research*, 36(12), 3152-3165.

Sim, T. K. T. (2007). The hub covering flow problem and the stochastic p-hub center problem. Iowa Research Online.URL: http://ir.uiowa.edu/etd/124. (access on Nov. 9, 2012)

Sim, T., et al. (2009). The stochastic p-hub center problem with service-level constraints. *Computers & Operations Research*, 36(12), 3166-3177.

Sinha, A. (2004). *Location, location, location and location: facility location incorporating demand uncertainty, logistic network design, product heterogeneity and competition*. PhD Thesis, Carnegie Mellon University; 2004

Skorin-Kapov, D. & Skorin-Kapov, J. (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research*, 73(3), 502-509.

Skorin-Kapov, D., et al. (1996). Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research*, 94(3), 582-593.

Smilowitz, K. & Daganzo, C. (2004). Cost modeling and design techniques for integrated package distribution systems. *Industrial Engineering and Management Science Working Paper, Northwestern University*, 04-06.

Smilowitz, K. R. & Daganzo, C. F. (2007). Continuum approximation techniques for the design of integrated package distribution systems. *Networks*, 50(3), 183-196.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing*, 21(1), 3-8.

Smith, K., et al. (1996). Neural versus traditional approaches to the location of interacting hub facilities. *Location Science*, 4(3), 155-171.

Sohn, J. & Park, S. (2000). The single allocation problem in the interacting three-hub network. *Networks*, 35(1), 17-25.

Song, Y. H. & Chou, C. S. V. (1999). Large-Scale Economic Dispatch by Artificial Ant Colony Search Algorithms. *Electric Machines & Power Systems*, 27(7), 679-690.

Stützle, T. (1999). Local search algorithms for combinatorial problems-analysis, algorithms and new applications. *DISKI-Dissertationen zur Künstliken Intelligenz.*

Snyder, L. V. (2006). Facility location under uncertainty: A review. *IIE Transactions*, 38(7), 547-564.

Snyder, L. V.&Daskin, M. S. (2006). Stochastic p-robust location problems. *IIE Transactions*, 38(11), 971-985.

Sohn, J. & Park, S. (1998). Efficient solution procedure and reduced size formulations for p-hub location problems. *European Journal of Operational Research*, 108(1), 118-126.

Sohn, J.&Park, S. (2000). The single allocation problem in the interacting three-hub network. *Networks*, 35(1), 17-25.

Sondergeld, L. & Voß, S. (1999). MIC-97: meta-heuristics international conference.

Southworth, F. & Peterson, B. E. (2000). Intermodal and international freight network modeling. *Transportation Research Part C: Emerging Technologies*, 8(1), 147-166.

Sridharan, R. (1993). A Lagrangian heuristic for the capacitated plant location problem with single source constraints. *European journal of operational research*, 66(3), 305-312.

Srinivas, M.&Patnaik, L. M. (1994). Genetic algorithms: A survey. *Computer*, 27(6), 17-26.

Taaffe, E. J., et al. (1996). *Geography of transportation.* Ed.2.Prentice Hall, New Jersey

Talbi, E. G. (2002). A taxonomy of hybrid metaheuristics. *Journal of heuristics*, 8(5), 541-564.

Talbi, E. G. (2009). *Metaheuristics: from design to implementation.* Wiley.

Tamir, A., et al. (2005). Conditional location of path and tree shaped facilities on trees. *Journal of algorithms*, 56(1), 50-75.

Tan, P. Z. & Kara, B. Y. (2007). A hub covering model for cargo delivery systems. *Networks*, 49(1), 28-39.

Thomadsen, T. & Larsen, J. (2007). A hub location problem with fully interconnected backbone and access networks. *Computers & operations research* , 34(8), 2520-2531.

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

Tonks, D. G. (2009). Validity and the design of market segments. *Journal of Marketing Management*, 25(3-4), 341-356.

Topcuoglu, H., et al. (2005). Solving the uncapacitated hub location problem using genetic algorithms. *Computers & Operations Research*, 32(4), 967-984.

Tuzun, D. & Burke, L. I. (1999). A two-phase tabu search approach to the location routing problem. *European Journal of Operational Research*, 116(1), 87-99.

United States International Trade Commission (USITC) (2004). *Express Delivery Services: Competitive Conditions Facing U.S.-based Firm in Foreign Markets.* USITC Publication 3678, U.S.A.

van Duin, R .& van Ham, H. (1998). A three-stage modeling approach for the design and organization of intermodal transportation services. Conference paper, *Systems, Man, and Cybernetics*, 1998. 1998 IEEE International Conference.

Van Landeghem, H. & Vanmaele, H. (2002). Robust planning: a new paradigm for demand chain planning. *Journal of Operations Management*, 20(6), 769-783.

Vazquez, M. & Whitley, D. (2000). A Comparison of Genetic Algorithms for the Static Job Shop Scheduling Problem. *Parallel Problem Solving from Nature PPSN VI*, 1917, 303-312.

Verter, V. & Dincer, M. C. (1995). Facility location and capacity acquisition: an integrated approach. *Naval Research Logistics (NRL)*, 42(8), 1141-1160.

Voss, S., et al. (1999). *Meta-heuristics: Advances and trends in local search paradigms for optimization.* Kluwer Academic Publishers.

Voss, S., et al. (1999). *Meta-heuristics: Advances and trends in local search paradigms for optimization.* Kluwer Academic Publishers.

Wagner, B. (2004). A note on" the latest arrival hub location problem". *Management science*, 50(12), 1751-1755.

Wagner, B. (2007). Model formulations for hub covering problems. *Journal of the Operational Research Society*, 59(7), 932-938.

Wagner, B. (2007). An exact solution procedure for a cluster hub location problem. *European journal of operational research*, 178(2), 391-401.

Wagner, B. (2008). A note on "Location of hubs in a competitive environment". *European Journal of Operational Research*, 184(1), 57-62.

Wang, M. & Zheng, C. (1997). Optimal remediation policy selection under general conditions. *Ground Water* ,35(5), 757-764.

Wang, Q., et al. (2003). Budget constrained location problem with opening and closing of facilities. *Computers & Operations Research*, 30(13), 2047-2069.

Wang, Z. N. & Liu, Z. L. (2000). *WTO and Post Service*. Beijing University of Posts and Telecommunications Press, Beijing.

Webb, M. H. J. (1968). Cost functions in the location of depots for multiple-delivery journeys. *Operations Reserch*, 311-320.

Wedel, M. & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations.* 8 Springer.

Weinstein, A. (2004). *Handbook of market segmentation: Strategic targeting for business and technology firms.* Routledge.

Weng, K. & Wang, Y. (2008). Evolutionary Algorithms for Multiple Allocation Hub Set Covering Problem. *Sensing and Control, 2008*, IEEE International Conference, 408-411.

Wasner, M. & Zäpfel, G. (2004). An integrated multi-depot hub-location vehicle routing model for network planning of parcel service. *International Journal of Production Economics*, 90(3), 403-419.

Wieberneit, N. (2008). Service network design for freight transportation: a review. *OR spectrum*, 30(1), 77-112.

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research*, 15(3), 317-337.

Wolpert, D. H. & Macready, W. G. (1997). No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on* ,1(1), 67-82.

Wu, T. H., et al. (2002). Heuristic solutions to multi-depot location-routing problems. *Computers & Operations Research*, 29(10), 1393-1415.

Yagiura, M. & Ibaraki, T. (2001). On metaheuristic algorithms for combinatorial optimization problems. *Systems and Computers in Japan*, 32(3), 33-55.

Yaman, H. (2011).Allocation strategies in hub networks. *European Journal of Operational Research*. 211(3), 442-451.

Yaman, H. & Carello, G. (2005). Solving the hub location problem with modular link capacities. *Computers & operations research*, 32(12), 3227-3245.

Yaman, H. (2006). Polyhedral analysis for the uncapacitated hub location problem with modular arc capacities. *SIAM Journal on Discrete Mathematics*, 19(2), 501.

Yaman, H., et al. (2007). The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B: Methodological*, 41(8), 906-919.

Yaman, H. (2009). The hierarchical hub median problem with single assignment. *Transportation Research Part B: Methodological*, 43(6), 643-658.

Yan, S., et al. (1995). Intermodal pricing using network flow techniques. *Transportation Research Part B: Methodological*, 29(3), 171-180.

Zäpfel, G. & Wasner, M. (2002). Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers. *International Journal of Production Economics*, 78(2), 207-220.

Zhao, H. et al. (2006). State-of-the-art and trend of development for Logistics in China. *China Logistics & Purchasing* , 9, 34-37.

Zhou, J. & Liu, B. (2007). Modeling capacitated location-allocation problem with fuzzy demands. *Computers & industrial engineering*, 53(3), 454-468.

Ziliaskopoulos, A. & Wardell, W. (2000). An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research* , 125(3), 486-502.

Declaration of Honor

I hereby confirm on my honor that the doctoral thesis submitted herewith is my own work.

All resources and aids that are used in my dissertation have been cited according to the rules for academic work and by means of footnotes or other precise indications of source.

The academic work has not been submitted to any other examination authority.

Hiermit melde ich mich bei meiner Ehre, zu bestätigen, dass die vorliegende Arbeit selbständig angefertigt habe.

Alle Ressourcen und Hilfsmittel, die in meiner Dissertation verwendet werden, sind nach den Regeln für die wissenschaftliche Arbeit und durch Fußnoten oder andere präzise Herkunftsangaben zitiert worden.

Die wissenschaftliche Arbeit hat keiner anderen Prüfungsbehörde vorgelegt worden.

Darmstadt, 30th January, 2013

Yida Xue

CURRICULUM VITAE

1. PERSONAL INFORMATION

- Name: Yida Xue
- Birthday: 30th December, 1983
- Gender: Female
- Nationality: P.R. China

2. Education

- 2009-present: Doctoral study at Section of Supply Chain and Network Management, Faculty of Laws and Economics, Technische Universität Darmstadt, Germany
- 2006-2008: Master of Science, School of Economics and Management, Tongji University, Shanghai, China
- 2002-2006: Bachelor of Science, School of Economics and Management, Tongji University, Shanghai, China