Shortest Paths and Multicommodity Network Flows

A Thesis
Presented to
The Academic Faculty

by

I-Lin Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
April 2003

Shortest Paths and Multicommodity Network Flows

Approved by:

Dr. Ellis L. Johnson, Committee Chair

Dr. Pamela H. Vance

Dr. George L. Nemhauser

Dr. Huei-Chuen Huang

Dr. Joel S. Sokol

Date Approved

To my Mom, wife, and son

for their unwavering love and support

ACKNOWLEDGEMENTS

First I want to thank Dr. Ellis Johnson for being my advisor and forseeing my potential in my dissertation topics that he suggested in 2000. In the beginning, I was a little reluctant and lacked confidence, since the topics were considered to be difficult to have new breakthrough. My progress turned out to be successful, because of Dr. Johnson's brilliant insight and continuous guidance over the years.

I would also like to thank Dr. Joel Sokol. Dr. Sokol was my former classmate at MIT, and was also on my committee. Although at MIT we didn't have a chance to chat too often, we do have frequent interaction at ISyE. He is very patient and has been most helpful in modifying my writings. He often suggests very clever advice which saves me time and inspires new ideas. I also acknowledge Dr. George Nemhauser, Dr. Pamela Vance, and Dr. Huei-Chuen Huang for having served on my committee. With their advice and supervision, my thesis became more complete and solid.

Special thanks to Dr. Gary Parker. Dr. Parker helped me find financial aid on my first day at GT. Also, he helped me to keep a 1/3 TA job when I took the comprehensive exam. I also thank Dr. Amy Pritchett for kindly supporting me in my first quarter. Without their support and Dr. Johnson's financial aid, I could not have earned my degree so easily. Thanks to Jennifer Harris for her adminstrative help over the years.

I thank Fujitsu Research Lab for giving me a foreign research position after MIT. Special thanks to Chujo san, Miyazaki san, Minoura san and Ogura san. Also I thank Yasuko & Satoru san, and my MIT senpai, Ding-Kuo and Hwa-Ping very much for the good times in Japan over the years.

Before coming to GT, I had an unhappy year at Stanford University. I have to thank Kien-Ming, Chih-Hung, and Li-Hsin for helping me out in that difficult time. Also thanks to Sheng-Pen, Tien-Huei, San-San, Che-Lin and Wei-Pang.

In my four and half years at ISyE, I have been helped by many good friends and classmates. Among them, I especially would like to thank Jean-Philippe. He is usually the first one I consulted with whenever I encountered research problems. I am in debt to Ricardo, Balaji, Dieter, Lisa, Brady, Andreea, Woojin, and Devang for their help. I also appreciated the help from Ivy Mok, Chun How Leung, Yen Soon Lee, and Dr. Loo Hay Lee in Singapore.

I am thankful for the many good people from Taiwan that helped me in Atlanta over the years, particulary Wei-Pin, Po-Long, Jenn-Fong, Chien-Tai & Chen-Chen, Tony & Wendy, James & Tina, Jack & Miawshian, Chih-Heng, Ivy & George, Chia-Chieh, Wen-Chih, Yen-Tai & Vicky, Po-Hsun and Cheng-Huang. I had a very great time with them.

I have to thank my Mom for her endless love and encouragement. Many thanks to my Mother and Father-in-law for taking care of my wife and son these years when I was away. Finally I want to thank my wife, Hsin-Yi, for supporting me to complete my degree and giving me a lovely son. Everytime when I would feel lonely or bored, I would watch our photos/videos, they always cheered me up and made me energetic again. I could not have done it without her unwavering love.

It has been 10 years, and I am always grateful for my Father in Heaven, for raising and educating me, which makes all of this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The shortest path problem is a classic and important combinatorial optimization problems. It often appears as a subproblem when solving difficult combinatorial problems like multicommodity network flow (MCNF) problems.

Most shortest path algorithms in the literature are either to compute the 1-ALL single source shortest path (SSSP) tree, or to compute the ALL-ALL all pairs shortest paths (APSP). However, most real world applications require only multiple pairs shortest paths (MPSP), where the shortest paths and distances between only some specific pairs of origin-destination nodes in a network are desired. The traditional single source shortest path (SSSP) and all pairs shortest paths (APSP) algorithms may do unnecessary computations to solve the MPSP problem.

We survey and summarize many shortest path algorithms, and discuss their pros and cons. We also investigate the Least Squares Primal-Dual method, a new LP algorithm that avoids degenerate pivots in each primal-dual iteration, for solving 1-1 and 1-ALL shortest path problems with nonnegative arc lengths, show its equivalence to the classic Dijkstra's algorithm, and compare it with the original primal-dual method.

We propose two new shortest path algorithms to save computational work when solving the MPSP problem. Our MPSP algorithms are especially suitable for applications with fixed network topology but changeable arc lengths. We discuss the theoretical details and complexity analyses. We test several implementations of our new MPSP algorithms extensively and compare them with many state-of-the-art SSSP algorithms for solving many families of artificially generated networks and a real Asia-Pacific flight network.

Our MPSP computational experiments show that there exists no ”killer” shortest path algorithm. Our algorithms have better performance for dense networks, but become worse for larger networks. Although they do not have best performance for the artificially generated graphs, they seem to be competitive for the real Aisa-Pacific flight network.

We provide an extensive survey on both the applications and solution methods for MCNF problems in this thesis. Among those methods, we investigate the combination of the primal-dual algorithm with the key path decomposition method. In particular, to efficiently solve the restricted primal problem (RPP) in each primal-dual iteration, we relax the nonnegativity constraints for some set of basic variables, which makes the relaxed RPP smaller and easier to solve since the convexity constraint will be implicitly maintained.

We implement our new primal-dual key path method (KEY), propose new methods to identify max step length, and suggest perturbation methods to avoid degenerate pivots and indefinite cycling problems caused by primal and dual degeneracy. We perform limited computational experiments to compare the running time of the generic primal-dual (PD) method, the Dantzig-Wolfe (DW) decomposition method, and the CPLEX LP solver that solves the node-arc form (NA) of the MCNF problems, with our method KEY. Observations from the computational experiments suggest directions for better DW implementation and possible remedies for improving PD and KEY.

CHAPTER I

INTRODUCTION TO MULTICOMMODITY NETWORK FLOW PROBLEMS

In this chapter, we first describe an air cargo flow problem in the Asia Pacific region. It is this problem that motivates our research. Because the problem can be modeled as a *multicommodity network flow* (MCNF) problem, we then introduce many MCNF related research topics and applications. Finally, we review different MCNF formulations and outline this thesis.

1.1 Asia Pacific air cargo system

Air cargo is defined as anything carried by aircraft other than mail, persons, and personal baggage. When an enterprise has facilities in its supply chain located in several different regions, efficient and reliable shipping via air is a crucial factor in its competitiveness and success. For example, many international electronics companies currently practice just-in-time manufacturing in which components are manufactured in China, Malaysia, Indonesia, Vietnam, or the Philippines, assembled in Taiwan, Japan or Singapore, and shipped to Europe and America.

The Asia Pacific region is expected to have the largest airfreight traffic in the next decade [177] due to the growth of Asia Pacific economies. In particular, China has drawn an increasing amount of investment from all over the world due to its cheap manpower and land. Many enterprises have moved their low-end manufacturing centers to China. Most China-made semifinished products are shipped to facilities in other Asian countries for key part assembly, and then sent out worldwide for sale.

The increasing need for faster air transportation drives the construction of larger and more efficient air facilities in the Asia Pacific region. New airports such as Chek Lap Kok airport (Hong Kong), Kansai airport (Osaka, Japan), Inchon airport (Seoul, Korea),

Table 1: Division of Asia Pacific into sub regions

Region	Countries in the Region
Central Asia	Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
South Asia	Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka
Northeast Asia	China, Hong Kong, Japan, Korea (Democratic People's Republic), Korea (Republic of), Macau, Mongolia, Russian Federation (East of Urals), Taiwan
Southeast Asia	Brunei Darussalam, Cambodia, Indonesia, Lao (Peoples' Democratic Republic), Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam
South Pacific	America Samoa, Australia, Christmas Island, Cocos (Keeling) Islands, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia, Nauru, New Caledonia, New Zealand, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Pitcairn, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, US Minor Outlying Islands, Vanuatu, Wallis and Futuna Islands

Bangkok airport (Thailand), and Pudong airport (Shanghai, Chian) have been constructed. Other airports such as Changi airport (Singapore), Narita airport (Tokyo, Japan), and Chiang-Kai-Shek airport (Taipei, Taiwan) have built new terminals or runways to be able to handle increased air traffic. A more thorough introduction to the air cargo system in the Asia Pacific region can be found in Bazaraa et al. [39].

This thesis is motivated by a study of the Asia Pacific air cargo system. We will describe an Asia Pacific flight network, and give a mathematical model for the air cargo system which corresponds to a multicommodity network flow (MCNF) model.

1.1.1 Asia Pacific flight network

The Asia Pacific region can be divided into five sub-regions as shown in Table 1.

Our research focuses on Northeast and Southeast Asia. Our Asia Pacific flight network (AP-NET) contains two types of nodes: center nodes that represent major cities (or airports) in the Northeast and Southeast Asia regions, and rim nodes that represent major cities in other regions. Arcs represent flight legs between center cities and center/rim cities. Based on data from the Freight Forecast [176], Air Cargo Annual [174], and Asia Pacific Air Transport Forecast [175] published by IATA, we selected 11 center countries (see Table 2) and 20 rim countries (see Table 3). The center countries are the Northeast and Southeast Asian countries that contribute the most intra and inter-regional air cargo traffic. The rim countries are those countries outside the region that have large air cargo traffic with the center countries.

Table 2: Center countries and cities of the AP filght network

center country	center city name
China	Beijing, Dalian, Guangzhou, Kunming, Qingdao, Shanghai, Shenyang, Shenzhen, Tianjin, Xi An, Xiamen
Hong Kong	Hong Kong
Indonesia	Denpasar Bali, Jakarta, Surabaya
Japan	Fukuoka, Hiroshima, Kagoshima, Komatsu, Nagoya, Niigata, Okayama, Okinawa, Osaka, Sapporo, Sendai, Tokyo
Korea	Cheju, Pusan, Seoul
Malaysia	Kota Kinabalu, Kuala Lumpur, Kuching, Langkawi, Penang
Philippines	Cebu, Manila, Subic Bay
Singapore	Singapore
Taiwan	Kaohsiung, Taipei
Thailand	Bangkok, Chiang Mai, Hat Yai, Phuket
Viet Nam	Hanoi, Ho Chi Minh City

Table 3: Rim countries and cities of the AP filght network

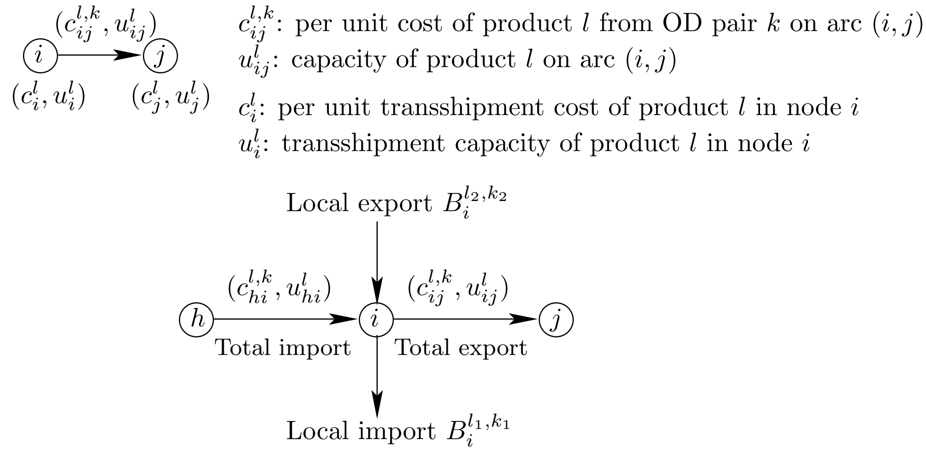
rim country	rim city name
Australia	Adelaide, Avalon, Brisbane, Cairns, Darwin, Melbourne, Perth, Sydney
Bahrain	Bahrain
Canada	Toronto, Vancouver
Denmark	Copenhagen
Finland	Helsinki
France	Paris
Germany	Frankfurt, Munich
Greece	Athens
India	Bangalore, Calcutta, Chennai, Delhi, Hyderabad, Mumbai
Italy	Milan, Rome
Netherlands	Amsterdam
New Zealand	Auckland, Christchurch
Pakistan	Islamabad, Karachi, Lahore
Russian Federation	Khabarovsk, Moscow, Novosibirsk, St Petersburg, Ulyanovsk
Sri Lanka	Colombo
Switzerland	Zurich
Turkey	Istanbul
United Arab Emirates	Abu Dhabi, Dubai, Sharjah
United Kingdom	London, Manchester
USA	Anchorage, Atlanta, Boston, Chicago, Dallas/Fort Worth, Detroit, Fairbanks, Honolulu, Houston, Kona, Las Vegas, Los Angeles, Memphis, Minneapolis/St Paul, New York, Oakland, Portland, San Francisco, Seattle, Washington

For our case study, we use the OAG MAX software package, which provides annual worldwide flight schedules for more than 850 passenger and cargo airlines, to build a monthly flight schedule of all jet non-stop operational flights between 48 center cities and 64 rim cities in September 2000. AP-NET contains 112 nodes and 1038 arcs, where each node represents a chosen city and each arc represents a chosen flight. Among the 1038 arcs, 480 arcs connect center cities to center cities, 277 arcs connect rim cities to center cities, and 281 arcs connect center cities to rim cities. We do not include the flights between rim cities in this model since we are only interested in the intra and inter-regional air cargo flows for the center nodes.

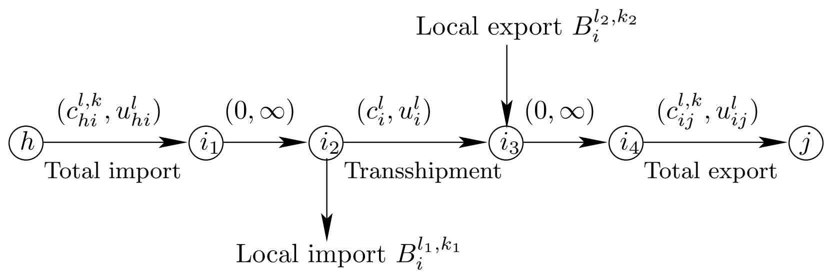
1.1.2 An air cargo mathematical model

The air cargo flow problem AP-NET is, in fact, a multicommodity network flow problem. In particular, given a monthly air cargo demand/supply data for both the center and rim nodes that have to be shipped via the flight legs (arcs) in AP-NET, then each commodity is a specific product for a specific origin-destination (OD) pair. Let $c_{i,j}^{l,k}$ denote the unit flow cost of shipping product l from origin s_k to destination t_k (i.e. OD pair k) on flight leg (i, j), and u_{ij}^l denote the capacity of shipping product l on flight leg (i, j). Each airport i has unit transshipment cost c_i^l and transshipment capacity u_i^l for product l (see Figure 1 (a)).

To eliminate the node cost and capacity, we can perform a node splitting procedure as in Figure 1 (b). In particular, suppose node i has to receive $B_i^{l_1,k_1}$ units of import cargo demand l_1 from node s_{k_1}, and send $B_i^{l_2,k_2}$ units of export cargo supply l_2 to node t_{k_2}. We can split node i into four nodes: an import node i_1 for receiving the import cargo, a demand node i_2 for receiving $B_{i_2}^{l_1,k_1}$ units of local import cargo, a supply node i_3 for sending $B_{i_3}^{l_2,k_2}$ units of local export cargo, and an export node i_4 for sending the export cargo. Suppose the original network has $|N|$ nodes and $|A|$ arcs. Then, the new network will have $4\,|N|$ nodes and $|A| + 3\,|N|$ arcs. The transshipment cost and capacity of node i become the cost and capacity of arc (i_2, i_3). Both arc (i_1, i_2) and arc (i_3, i_4) have zero costs and unbounded capacities..



(a) Original network for product l



(b) Transformed network for product l

Figure 1: Network transformation to remove node cost and capacity

In this way, the minimum cost cargo routing problem becomes a minimum cost multicommodity network flow problem. It seeks optimal air cargo routings to minimize the total shipping cost of satisfying the OD demands while not exceeding the airport or flight capacities. Each commodity is defined as a specific product to be shipped from an origin to a destination.

If we have all of the parameters (OD demand data for each node, unit shipping and transshipment costs for each commodity on each flight leg and airport, and the capacity of each flight leg and airport), this MCNF problem can be solved to optimality. The sensitivity analysis can give us insight and suggest ways to improve the current Asia Pacific air cargo system. For example, by perturbing the capacity for a specific flight leg or airport, we can determine whether adding more flights or enlarging an airport cargo terminal is profitable, which flight legs are more important, and whether current practices for certain airline or airport are efficient. Similar analyses can be done to determine whether adding a new origin-destination itinerary is worthwhile, or whether increasing the charge on a certain OD itinerary is profitable.

However, it is difficult to determine those parameters. The flight capacity may be estimated by adding the freight volumes of each aircraft on a certain flight leg. The unit shipping and transshipping cost or revenue is difficult to determine, since it is a function of several variables such as time, distance, product type and OD itinerary (including the flight legs and airports passed). The transshipment capacity for a cargo terminal is also difficult to estimate since it is variable and dynamic. We may first treat it as unbounded and use the optimal transshipment flow to judge the utility of cargo terminals. Correctly estimating these parameters is itself a challenging research topic. This thesis, on the other hand, focuses on how to solve the MCNF problem more efficiently, given all these parameters.

1.2 MCNF problem description

The multicommodity network flow problem is defined over a network where more than one commodity needs to be shipped from specific origin nodes to destination nodes while not violating the capacity constraints associated with the arcs. It extends the *single commodity network flow* (SCNF) problem in a sense that if we disregard the *bundle constraints*, the arc capacity constraints that tie together flows of different commodities passing through the same arc, a MCNF problem can be viewed as several independent SCNF problems.

In general, there are three major MCNF problems in the literature: the *max MCNF* problem, the *max-concurrent flow* problem, and the *min-cost MCNF* problem. The max MCNF problem is to maximize the sum of flows for all commodities between their respective origins and destinations. The max-concurrent flow problem is a special variant of the max MCNF problem which maximizes the fraction (throughput) of the satisfied demands for all commodities. In other words, the max concurrent flow problem maximizes the fraction z for which the min-cost the MCNF problem is feasible if all the demands are multiplied by z. The min-cost MCNF problem is to find the flow assignment satisfying the demands of all commodities with minimum cost without violating the capacity constraints on all arcs.

$|N| = 5$, $|A| = 10$, $|K| = 3$

K^1: $(s^1, t^1) = (1, 4)$, $B^1 = 6$
K^2: $(s^2, t^2) = (5, 3)$, $B^2 = 4$
K^3: $(s^3, t^3) = (3, 1)$, $B^3 = 8$

origins, destinations, and demands
of three commodities

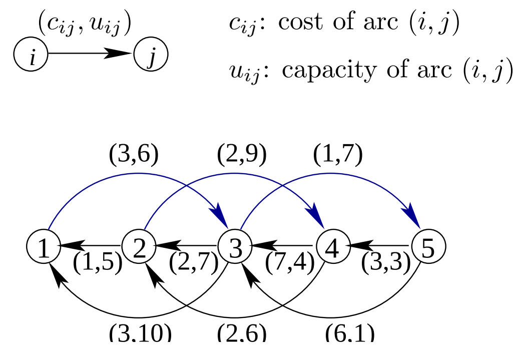


Figure 2: A MCNF example problem

This dissertation will focus on min-cost MCNF problem. Hereafter, when we refer to the MCNF problem, we mean the min-cost MCNF problem.

The existence of the bundle constraints makes MCNF problems much more difficult than SCNF problems. For example, many SCNF algorithms exploit the single commodity flow property in which flows in opposition direction on an arc could be canceled out. In MCNF problems, flows do not cancel if they are different commodities. The max-flow min-cut theorem of Ford and Fulkerson [116] in SCNF problems guarantees that the maximum flow is equal to the minimum cut. Furthermore, with integral arc capacities and node demands, the maximum flow is guaranteed to be integral. None of these properties can be extended to MCNF, except for several special planar graphs [255, 254, 224, 250], or two-commodity flows with even integral demands and capacities [170, 276, 278]. The *total unimodularity* of the constraint matrix in the SCNF LP formulation guarantees integral optimal flows in the cases of integral node supplies/demands and arc capacities. This integrality property also can not be extended to MCNF LP formulations. Solving the general integer MCNF problem is NP-complete [194]. In fact, even solving the two-commodity integral flow problem is NP-complete [107].

Since 1960s, MCNF has motivated many research topics. For example, *column generation* by Ford and Fulkerson [115] was originally designed to solve max MCNF problems, and is still a common technique for solving large-scale LP problems. Seymour [286] proposes the *max-flow min-cut matroid* and several important matroid theorems by studying multicommodity flow. The nonintegrality property of MCNF spurs much research in integral

$$\widetilde{N} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Node-arc incidence matrix $\widetilde{N}$

$$\begin{bmatrix} \widetilde{N} & & \\ & \widetilde{N} & \\ & & \widetilde{N} \\ I & I & I \end{bmatrix} X \begin{array}{l} = b_1 = [\ 6\ 0\ 0\ \text{-}6\ 0\ ]^T \\ = b_2 = [\ 0\ 0\ \text{-}4\ 0\ 4\ ]^T \\ = b_3 = [\ \text{-}8\ 0\ 8\ 0\ 0\ ]^T \\ \leq u = [\ 6\ 5\ 9\ 10\ 7\ 7\ 6\ 4\ 1\ 3\ ]^T \end{array}$$

$$X = [\ x^1\ x^2\ x^3\ ]^T\ ,\ x^k \in R_+^{|A|}\ \forall k \in K$$

MCNF constraints

Figure 3: Block angular structure of the example problem in Figure 2

LP related to matroids, algorithmic complexity, and polyhedral combinatorics (see [162] for details). The block-angular constraint structure (see Figure 3) of the MCNF constraints serves as a best practice for decomposition techniques such as *Dantzig-Wolfe decomposition* [88] and *Benders decomposition* [42], and basis partitioning methods such as *generalized upper bounding* [213, 167] and *key variables* [273].

1.3 Applications

MCNF models arise in many real-world applications. Most of them are network routing and network design problems.

1.3.1 Network routing

1.3.1.1 Message routing in Telecommunication:

Consider each requested OD pair to be a commodity. The problem is to find a min-cost flow routing for all demands of requested OD pairs while satisfying the arc capacities. This appears often in telecommunication. For example, message routing for many OD pairs [36, 237], packet routing on virtual circuit data network [221] (all of the packets in a session are transmitted over exactly one path between the OD to minimize the average number of packets in the network), or routing on a ring network [288] (any cycle is of length n), are

MCNF problems.

1.3.1.2 *Scheduling and routing in Logistics and Transportation:*

In logistics or transportation problems, commodities may be objects such as products, cargo, or even personnel. The commodity scheduling and routing problem is often modeled as a MCNF problem on a time-space network where a commodity may be a tanker [41], aircraft [166], crew [60], rail freight [16, 78, 210], or Less-than Truck-Load (LTL) shipment [109, 38].

Golden [144] gives a MCNF model for port planning that seeks optimal simultaneous routing where commodities are export/import cargo, nodes are foreign ports (as origins), domestic hinterlands (as destinations) and US ports (as transshipment nodes), and arcs are possible routes for cargo traffic. Similar problems appear in grain shipment networks [9, 31].

A disaster relief management problem is formulated as a multicommodity multimodal network flow problem with time windows by Haghani and Oh [164] where commodities (food, medical supplies, machinery, and personnel) from different areas are to be shipped via many modes of transportation (car, ship, helicopter,...,etc.) in the most efficient manner to minimize the loss of life and maximize the efficiency of the rescue operations.

1.3.1.3 *Production scheduling and planning:*

Jewell [180] solves a warehousing and distribution problem for seasonal products by formulating it as a MCNF model where each time period is a transshipment node, one dummy source and one dummy sink node exist for each product, and arcs connect from source nodes to transshipment nodes, earlier transshipment nodes to later transshipment nodes, and transshipment nodes to sink nodes. Commodities are products to be shipped from sources to sinks with minimum cost.

D'Amours et al. [84] solve a planning and scheduling problem in a Symbiotic Manufacturing Network (SMN) for a multiproduct order. A broker receives an order of products in which different parts of the products may be manufactured by different manufacturing firms, stored by some storage firms, and shipped by a few transportation firms between manufacturing firms, storage firms and customers. The problem is to design a planning and scheduling bidding scheme for the broker to make decisions on when the bids should be assigned and who they should be assigned to, such that the total cost is minimized. They propose a MCNF model where each firm (manufacturing or storage) at each period represents a node, an arc is either a possible transportation link or possible manufacturing (storage) decision, and a commodity represents an order for different product.

Aggarwal et al. [1] use a MCNF model to solve an equipment replacement problem which determines the time and amount of the old equipments to be replaced to minimize the cost.

1.3.1.4 Other routing:

- **VLSI design** [61, 279, 270, 8]: Global routing in VLSI design can be modeled as an origin-destination MCNF problem where nodes represent collections of terminals to be wired, arcs correspond to the channels through which the wires run, and commodities are OD pairs to be routed.

- **Racial balancing of schools** [77]: Given the race distribution of students in each community, and the location, capacity and the ethnic composition requirements of each school, the problem is to seek an assignment of students to schools so that the ethnic composition at each school is satisfied, no student travels more than a specified number of minutes per day, and the total travel time for students to school is minimized. This is exactly a MCNF problem in the sense that students of the same race represent the same commodity.

- **Caching and prefetching problems in disk systems** [6, 7]: Since the speed of computer processors are much faster than memory access, caching and prefetching are common techniques used in modern computer disk systems to improve the performance of their memory systems. In prefetching, memory blocks are loaded from the disk (slow memory) into the cache (fast memory) before actual references to the blocks, so the waiting time is reduced in accessing memory from the disk. Caching tries to keep actively referenced memory blocks in fast memory. At any time, at most one fetch operation is executed. Albers and Witt [7] model this problem as a min-cost

MCNF problem that decides when to initiate a prefetch, and what blocks to fetch and evict.

- **Traffic equilibrium** [215, 110]: The traffic equilibrium law proposed by Wardrop [303] states that at equilibrium, for each origin-destination pair, the travel times on all routes used are equal, and are less than the travel times on all nonused routes. Given the total demand for each OD pair, the equilibrium traffic assignment problem is to predict the traffic flow pattern on each arc for each OD pair that follows Wardrop's equilibrium law. It can be formulated as a nonlinear MCNF problem where the bundle constraints are eliminated but the nonlinear objective function is designed to capture the flow congestion effects.

- **Graph Partitioning** [285]: The graph partitioning problem is to partition a set of nodes of a graph into disjoint subsets of a given maximal size such that the number of arcs with endpoints in different subsets is minimized. It is an NP-hard problem. Based on a lower bounding method for the graph bisection problem, which corresponds to a MCNF problem, Sensen [285] proposes three linear MCNF models to obtain lower bounds for the graph partitioning problem. He then uses branch-and-bound to compute the exact solution.

  Similarly, many NP-hard problems such as min-cut linear arrangement, crossing number, minimum feedback arc set, minimum 2D area layout, and optimal matrix arrangement for nested disection can be approximately solved by MCNF algorithms [217, 204, 206, 219].

1.3.2 Network design

Given a graph G, a set of commodities K to be routed according to known demands, and a set of facilities L that can be installed on each arc, the *capacitated network design* problem is to route flows and install facilities at minimum cost. This is a MCNF problem which involves flow conservation and bundle constraints plus some side constraints related to the installation of the new facilities. The objective function may be nonlinear or general discontinuous step-increasing [124].

Problems such as the design of a network where the maximum load on any edge is minimized [51] or the service quality and survivability constraints are met with minimum cost of additional switches/transport pipes in ATM networks [53] appear often in the telecommunication literature. Bienstock et al. use metric inequalities, aggregated formulations [50] and facet-defining inequalities [52] to solve capacitated survivable network design problems. Gouveia [152] discusses MCNF formulations for a specialized *terminal layout problem* which seeks a minimum spanning tree with hop constraints (a limit on the number of hops (links) between the computer center and any terminal in the network).

Maurras et al. [232, 163] study network design problems with jump constraints (i.e., each path has no more than a fixed number of arcs). Girard and Sansò [133] show that the network designed using a MCNF model significantly improves the robustness of the heuristic solutions at a small cost increase. Gendron et al. [130] write a comprehensive survey paper in multicommodity capacitated network design.

Similar problems also appear in transportation networks, such as locating vehicle depots in a freight transportation system so that the client demands for empty vehicles are satisfied while the depot opening operating costs and other transportation costs are minimized. Crainic et al. [80] have solved this problem by various methods such as branch-and-bound [82] and its parallelization [129, 56], dual-ascent [81], and tabu-search [83]. Crainic also writes a survey paper [79] about service network design in freight transportation. In fact, these problems are facility location problems. Geoffrion and Graves [132] model a distribution system design problem as a fixed charge MCNF problem. More MCNF models for facility location are surveyed by Aikens [5].

1.4 Formulations

Let N denote the set of all nodes in G, A the set of all arcs, and K the set of all commodities. For commodity k with origin s_k and destination t_k, c^k_{ij} represents its per unit flow cost on arc (i,j) and x^k_{ij} the flow on arc (i,j). Let b^k_i be the supply/demand at node i, and B^k be the total demand units of commodity k. Let u_{ij} be the arc capacity on arc (i,j). Without loss of generality, we assume each unit of each commodity consumes one unit of capacity from each arc on which it flows.

1.4.1 Node-Arc form

The node-arc form of MCNF problem is a direct extension of the conventional SCNF formulation. It can be formulated as follows:

$$\min \sum_{k \in K} \sum_{(i,j) \in A} c_{ij}^k x_{ij}^k = Z^*(x) \tag{Node-Arc}$$

$$s.t. \sum_{(i,j) \in A} x_{ij}^k - \sum_{(j,i) \in A} x_{ji}^k = b_i^k \quad \forall i \in N, \forall k \in K \tag{1.1}$$

$$\sum_{k \in K} x_{ij}^k \leq u_{ij} \quad \forall (i,j) \in A \text{ (bundle constraints)} \tag{1.2}$$

$$x_{ij}^k \geq 0 \quad \forall (i,j) \in A, \forall k \in K$$

where $b_i^k = B^k$ if $i = s_k$, $b_i^k = -B^k$ if $i = t_k$, and $b_i^k = 0$ if $i \in N \backslash \{s_k, t_k\}$. This formulation has $|K||A|$ variables and $|N||K| + |A|$ nontrivial constraints.

1.4.2 Arc-Path form

First suggested by Tomlin [297], the arc-path form of the min-cost MCNF problem is based on the fact that any network flow solution can be decomposed into path flows and cycle flows. Under the assumption that no cycle has negative cost, any arc flow vector can be expressed optimally by simple path flows.

For commodity k, let P^k denote the set of all possible simple paths from s_k to t_k, f_p the units of flow on path $p \in P^k$, and PC_p^c the cost of path p using c_{ij}^k as the arc costs. δ_a^p is a binary indicator which equals to 1 if path p passes through arc a, and 0 otherwise. It can be formulated as follows:

$$\min \sum_{k \in K} \sum_{p \in P^k} PC_p^c f_p = Z^*(f) \tag{Arc-Path}$$

$$s.t. \quad \sum_{p \in P^k} f_p = 1 \quad \forall k \in K \tag{1.3}$$

$$\sum_{k \in K} \sum_{p \in P^k} (B^k \delta_a^p) f_p \leq u_a \quad \forall a \in A \text{ (bundle constraints)} \tag{1.4}$$

$$f_p \geq 0 \quad \forall p \in P^k, \forall k \in K$$

Inequalities (1.4) are the bundle constraints, and (1.3) are the convexity constraints which force the optimal solution for each commodity k to be a convex combination of some simple paths in P^k. Since we only consider the LP MCNF problem in this dissertation, the optimal solution can be fractional. For the binary MCNF problem where each commodity can only be shipped on one path, f_p will be be binary variables.

This formulation has $\sum_{k \in K} |P^k|$ variables and $|K| + |A|$ nontrivial constraints.

1.4.3 Comparison of formulations

When commodities are OD pairs, $|K|$ may be $O(|N|^2)$ in the worst case. In such case, the node-arc form may have $O(|N|^3)$ constraints which make its computations more difficult and memory management less efficient.

The arc-path formulation, on the other hand, has at most $O(|N|^2)$ constraints but exponentially many variables. The problem caused by a huge number of variables can be resolved by column-generation techniques. In particular, when using the revised simplex method to solve the arc-path form, at most $|K| + |A|$ of the $\sum_{k \in K} |P^k|$ variables are positive in any basic solution. In each iteration of the revised simplex method, a shortest path subproblem for each commodity can be solved to find a new path for a simplex pivot-in operation if it corresponds to a nonbasic column with negative reduced cost.

Dantzig-Wolfe (DW) decomposition is a common technique used to solve problems with block-angular structure. The algorithm decomposes the problem into two smaller sets of problems, a *restricted master problem* (RMP) and k subproblems, one for each commodity. It first solves the k subproblems and uses their basic columns to solve the RMP to optimality. Then it uses the dual variables of the RMP to solve the k subproblems, which produce nonbasic columns with negative reduced cost to be added to the RMP. The procedures are repeated until no more profitable columns are generated for the RMP. More details about DW will be discussed in Section 2.3.2.

Jones et al. [187] investigate the impact of formulation on Dantzig-Wolfe decomposition for the MCNF problem. Three formulations: origin-destination specific (ODS), destination specific (DS), and product specific (PS) are compared in Table 4 where shortest path, shortest path tree, and min-cost network flow problems are the subproblems for ODS, DS, and PS respectively.

Table 4: Comparison of three DW formulations [187]

subproblem*	$ODS \prec DS \prec PS$
RMP size	$ODS > DS > PS$
RMP sparsity**	$ODS \succ DS \succ PS$
number of master iterations	$ODS < DS < PS$
number of commodities	$ODS \gg DS > PS$
convergence rate	$ODS > DS > PS$
*: $A \succ B$ means A is harder than B ; **: $A \succ B$ means A is sparser than B	

Although the ODS form has a larger RMP and more subproblems (due to more commodities), it has a sparser RMP, easier subproblems, and better extreme points produced which contribute to faster convergence than other forms.

Knowing this fact, our solution methods to the min-cost ODMCNF problem will be path-based techniques.

1.5 Contributions and thesis outline

We consider the following to be the major contributions of this thesis:

- We survey and summarize MCNF problems and algorithms that have appeared in the last two decades.
- We survey and summarize shortest path algorithms that have appeared in the last five decades.
- We observe the connection between a new shortest path algorithm and the well-known Dijkstra's algorithm.
- We propose three new multiple pairs shortest path algorithms, discuss their theoretical properties, and analyze their computational performance.
- We give two new MCNF algorithms, and compare their computational performance with two standard MCNF algorithms.

This thesis has six chapters. Chapter 1 discusses MCNF applications and their formulations. Chapter 2 surveys MCNF solution techniques from the literature. Since our approaches require us to solve many *multiple pairs shortest paths* (MPSP) problems, we will propose new efficient MPSP algorithms. Chapter 3 first reviews the shortest path solution techniques that have appeared in the last five decades, and then introduces a new shortest path method which uses the *nonnegative least squares* (NNLS) technique in the *primal-dual* (PD) algorithmic framework, and discusses its relation to the well-known Dijkstra's algorithm. Chapter 4 discusses the theoretical details of our new MPSP algorithms. Chapter 5 presents the computational performance of our new MPSP algorithms. Finally, Chapter 6 illustrates our primal-dual MCNF algorithms, analyzes their computational performance, and suggests new techniques to improve the algorithms. Chapter 7 concludes this thesis and suggests future research.