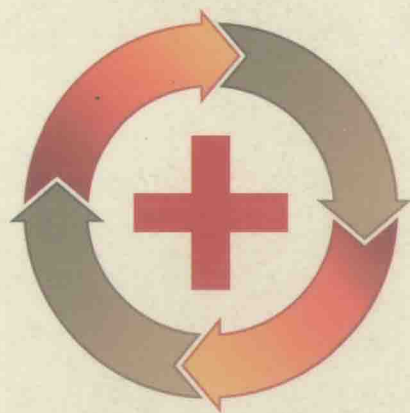


质量与效率的平衡术

——以医疗与安检服务运营为例

王晓芳 著



**BALANCING SERVICE
QUALITY & EFFICIENCY
IN HEALTHCARE AND SECURITY SYSTEMS**

Xiaofang Wang



知识产权出版社
INTELLECTUAL PROPERTY PUBLISHING HOUSE

质量与效率的平衡术

——以医疗与安检服务运营为例

Balancing Service Quality and Efficiency
in Healthcare and Security Systems



知识产权出版社

INTELLECTUAL PROPERTY PUBLISHING HOUSE

内容提要

近年来,流行性疾病、自然或人为灾害频繁暴发,对各国的卫生防御体系和安全防御体系带来了巨大挑战。医疗呼叫中心在应对突发紧急情况分诊分流方面具有有效分配医疗资源的巨大潜力,安检系统(包括海关、机场、签证等)也正积极探索如何采用运营管理有效地准备和响应突发事件,并尽可能减少对公众正常生活的负面影响。

在此背景之下,本书深入研究了这类特殊服务系统的运营特点和决策优化,重点分析了其特有的质量/效率均衡关系。

本书将最前沿的服务运营研究成果与实证发现相结合,通过严密的数学分析提供详细、系统化的管理方法,具有很强的现实指导意义。

责任编辑:江宜玲

责任出版:卢运霞

图书在版编目(CIP)数据

质量与效率的平衡术:以医疗与安检服务运营为例:英文/王晓芳著. —北京:知识产权出版社,2011.6

ISBN 978-7-5130-0607-1

I. ①质… II. ①王… III. ①医疗卫生服务—运营管理—研究—英文②安全检查—运营管理—研究—英文 IV. ①F719②X924

中国版本图书馆CIP数据核字(2011)第104550号

质量与效率的平衡术——以医疗与安检服务运营为例

ZHILIANG YU XIAOLÜ DE PINGHENGSHU——YI YILIAO YU ANJIAN FUWU
YUNYING WEILI

王晓芳 著

出版发行: 知识产权出版社

社 址: 北京市海淀区马甸南村1号

网 址: <http://www.ipph.cn>

发行电话: 010-82000860 转 8101/8102

责编电话: 010-82000860 转 8339

印 刷: 北京富生印刷厂

开 本: 880mm×1230mm 1/32

版 次: 2011年8月第1版

字 数: 144千字

ISBN 978-7-5130-0607-1/F·425 (3509)

邮 编: 100088

邮 箱: hjb@cnipr.com

传 真: 010-82000507/82000893

责编邮箱: jiangyiling@cnipr.com

经 销: 新华书店及相关销售网点

印 张: 5.5

印 次: 2011年8月第1次印刷

定 价: 28.00元

出版版权专有 侵权必究

如有印装质量问题,本社负责调换。

To

**my teachers, colleagues and professionals
who shared with me their knowledge**

Preface

In recent years, frequent epidemic outbreaks and natural or man-made disasters are challenging the disease control/prevention and security systems in many countries. Facing this challenge, a type of healthcare service center has shown its great potentials in providing telephone triage to help ensure that medical care is delivered at the appropriate level in response to emergency events. Similarly, in security screening including visa issuance, cargo inspection, and airport security screening, increased attention has been paid to operational techniques to better respond to natural and man-made hazards and mitigate the negative impact on people's life.

Motivated by real applications, this book studies an important aspect of service operations in the above special areas—the quality and efficiency tradeoff. High speed/efficiency is important but not at the sacrifice of quality. In addition, considering the management uncertainty and complexity involved in these highly human-centric contexts, this book focuses on modeling and analyzing the service provider's autonomy in determining service time and service depth, as well as customers' autonomy in service choice, especially the impact of the equilibrium behavior in service design.

The goal of this book is to provide analytical frameworks for modeling this aspect, and understanding its effect on costs and revenue in service operations management. To accomplish this goal, three management problems are formulated and analyzed in healthcare systems, security systems, and scheduling, respectively. The novelty of this book,

compared to previous studies, is that it highlights and analyzes in detail the quality and efficiency tradeoff widely existing in healthcare and security. Results generated in this book are expected to provide systematic guidance and valuable insights for current practice.

This book is mostly based on my Ph. D. thesis work at the Tepper School of Business, Carnegie Mellon University and other work after I join the faculty group at the School of Business, Renmin University of China. I am grateful to my Ph. D. advisors Professor Alan Scheller-Wolf and Professor Stephen F. Smith at Carnegie Mellon University for their constant support and guidance throughout my graduate program. I would also like to express my gratitude to Professor Laurens Debo at University of Chicago, who initiated a challenging but interesting project in healthcare and guided me to finish the research work in healthcare. The knowledge and ideas in the security areas have greatly benefited from my collaboration with Professor Jun Zhuang at University of Buffalo. The knowledge and ideas in healthcare practice have benefited from Professor Martin Gaynor at Carnegie Mellon University who initiated my contacts with the healthcare insurance companies. Some of my ideas and practical knowledge in healthcare originate from my interactions with the people at Children's Hospital of Pittsburgh. I would particularly like to thank Dr. Richard Saladino and Ms. Deborah Lesniak for inviting me to join their project team and helping me reach the right information sources. I also thank Jennifer Iagnemma and Tom Gentle for their selfless help. I thank my colleagues at Renmin University of China and my family. Without their care and help, I cannot enjoy the process of writing this book. Finally, I thank the funding support by the National Natural Science Foundation of China (No. 71001102).

Xiaofang Wang

Contents

| | | |
|------------------|--|----|
| Chapter 1 | Introduction | 1 |
| Chapter 2 | Diagnostic Service Centers | 4 |
| 2.1 | Literature Review | 10 |
| 2.2 | The Model | 14 |
| 2.3 | General Analysis | 21 |
| 2.4 | Analysis of the Symmetric Case | 27 |
| 2.5 | Analysis of the Asymmetric Case | 32 |
| 2.6 | Numerical Case Studies and Extensions | 34 |
| 2.6.1 | Parameters and Base Case | 35 |
| 2.6.2 | Changing the HO's error costs | 35 |
| 2.6.3 | Changing the patients' error costs | 37 |
| 2.6.4 | Convex waiting costs | 38 |
| 2.7 | Relevance of Integrated Model and Analysis to Practice | 41 |
| 2.8 | Discussion and Conclusion | 48 |
| Chapter 3 | Security Screening | 52 |
| 3.1 | The Model | 56 |
| 3.1.1 | Model Setup | 58 |
| 3.1.2 | Definition of Equilibrium | 61 |
| 3.2 | Potential Applicants' Best Responses | 63 |
| 3.2.1 | Bad Potential Applicants' Best Responses | 63 |
| 3.2.2 | Good Potential Applicants' Best Responses | 63 |
| 3.3 | Optimal Screening Policies | 67 |
| 3.3.1 | Non-discriminatory Screening Policy | 67 |
| 3.3.2 | Discriminatory Screening Policy | 71 |

| | | |
|------------------|---|-----------|
| 3.4 | Numerical Experiments | 73 |
| 3.4.1 | Changing the Approver's Reward from Admitting a Good Applicant R | 74 |
| 3.4.2 | Changing the Parameter of Bad Potential Applicants r_b | 75 |
| 3.4.3 | Changing the Arrival Rate of All Potential Applicants Λ | 77 |
| 3.4.4 | Changing Service Capacity μ | 79 |
| 3.5 | Extensions of Our Model | 80 |
| 3.5.1 | Imperfect Screening | 80 |
| 3.5.2 | Non-zero Abandon Rate for Good Applicants | 88 |
| 3.5.3 | Non-zero Waiting Costs for Bad Applicants | 89 |
| 3.6 | Conclusion and Future Research Directions | 91 |
| Chapter 4 | High Quality Schedules | 93 |
| 4.1 | Literature Review | 97 |
| 4.1.1 | Time-cost Tradeoff Problems (Project Crashing) | 98 |
| 4.1.2 | Anytime Scheduling | 98 |
| 4.1.3 | Hybrid Scheduling Techniques | 98 |
| 4.1.4 | Temporal Preference Networks | 100 |
| 4.1.5 | Service Operations Management | 100 |
| 4.2 | Problem Formulation | 101 |
| 4.3 | Problem Solving via Constraint-based Algorithms | 105 |
| 4.3.1 | The Integrated (Quality First) Hybrid Scheme | 106 |
| 4.3.2 | The Partitioned (Resource First) Hybrid Scheme; Solve-and-Maximize | 113 |
| 4.4 | Evaluation on Performance of Heuristics | 121 |
| 4.4.1 | Ratio vs. Slope | 124 |
| 4.4.2 | Exact Lookahead or Not? | 128 |
| 4.4.3 | Impact of Longest Peak First (LPF) | 132 |

| | | |
|---------------------|---|-----|
| 4.4.4 | Summary | 133 |
| 4.5 | Evaluation of the Two Constraint-based Methods | 133 |
| 4.5.1 | Integrated vs. Partitioned | 135 |
| 4.5.2 | A Further Analysis of Temporal Flexibility | 138 |
| 4.6 | Conclusions and Discussion | 139 |
| 4.6.1 | Other Quality Profiles | 140 |
| 4.6.2 | Other Scheduling Settings | 141 |
| Chapter 5 | Conclusions and Future Research Directions | 146 |
| 5.1 | Summary of Contributions | 146 |
| 5.2 | Extensions and Future Work | 149 |
| Bibliography | | 152 |

List of Figures

| | | |
|------|--|-----|
| 2.1 | Optimal error rates and staffing level vs. asymmetry in the HO's error costs | 36 |
| 2.2 | Optimal error rates and staffing level vs. asymmetry in the patient's error costs | 37 |
| 2.3 | Optimal error rates and staffing levels when waiting costs are convex | 40 |
| 3.1 | The approval process with strategic potential applicants and congestion | 57 |
| 3.2 | Potential applicants' best responses as functions of Φ_1 and Φ_2 | 64 |
| 3.3 | Potential applicants' best responses and approver's payoff using the non-discriminatory screening policy | 70 |
| 3.4 | Approver's payoffs as functions of Φ_1 and Φ_2 | 72 |
| 3.5 | Sensitivity analysis for the approver's gain R | 74 |
| 3.6 | Sensitivity analysis for r_b | 76 |
| 3.7 | Sensitivity analysis for Λ | 78 |
| 3.8 | Sensitivity analysis for μ | 80 |
| 3.9 | Sensitivity analysis for e_b | 87 |
| 3.10 | Sensitivity analysis for e_g | 87 |
| 3.11 | Sensitivity analysis for γ | 89 |
| 3.12 | Sensitivity analysis for b_w | 90 |
| 4.1 | The precedence constraint posting framework | 107 |
| 4.2 | An example for quality loss estimation | 112 |
| 4.3 | Types of scheduling solutions | 114 |

| | | |
|------|---|-----|
| 4.4 | Chaining-form schedule linking the activities into four chains | 116 |
| 4.5 | A <i>POS</i> (left) vs. a more flexible <i>POS</i> (right) generated by the <i>minimizing the interdependencies</i> heuristic | 120 |
| 4.6 | Steps in the partitioned hybrid algorithm | 121 |
| 4.7 | Comparison between slope and ratio | 125 |
| 4.8 | A final schedule generated by Global-Slope | 127 |
| 4.9 | Different quality loss computation methods | 129 |
| 4.10 | Comparison on number of problem solved between the integrated approach and the partitioned approach | 136 |
| 4.11 | Comparison on quality between the integrated approach and the partitioned approach | 137 |
| 4.12 | Comparison on number of posted constraints between the integrated approach and the partitioned approach | 137 |
| 4.13 | The effects of temporal flexibility on the quality solution | 139 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Error costs for the patient and HO as a function of the patient's pathology and selected treatment | 15 |
| 2.2 | Summary of notations | 21 |
| 2.3 | Sensitivity analysis for symmetric case when $c_w > 0$ | 29 |
| 2.4 | Impact of asymmetry (increasing C_A to $C_A + \delta$) | 33 |
| 2.5 | Parameter settings used for the numerical experiments | 35 |
| 2.6 | Protocols and their implementations | 45 |
| 3.1 | Major notations | 62 |
| 3.2 | The optimal non-discriminatory screening policy | 68 |
| 3.3 | The optimal non-discriminatory (imperfect) screening policy ($e_b > \frac{c_b}{c_b + r_b}$ and $\hat{p}_b = 1$) | 84 |
| 3.4 | The optimal non-discriminatory (imperfect) screening policy ($e_b \leq \frac{c_b}{c_b + r_b}$) | 85 |
| 4.1 | Constraint posting heuristics | 109 |
| 4.2 | Overall performance results | 123 |
| 4.3 | Data for the activities in the sequence | 126 |
| 4.4 | Precedence relations in a resulting schedule | 129 |
| 4.5 | Experimental results | 135 |
| 4.6 | Quality values using <i>simple chaining</i> vs. <i>iterative chaining</i> | 138 |

Chapter 1

Introduction

The continually escalating expenditures for healthcare and frequent epidemic outbreaks have motivated this industry to seek guidance from Operations Management (OM). This has not always been easy: A significant challenge of applying operations management theory to healthcare is that in contrast to manufacturing or other service systems, healthcare service delivery is a human-centric and high-risk process. Thus well-established operations management theories need to be further studied in order to address typical concerns in healthcare such as quality, safety, efficiency, staffing and patient satisfaction.

Similarly, natural or man-made disasters have challenged the current security systems in many countries. Specially, in security screening including visa issuance, cargo inspection, and airport security screening, increased attention has been paid to operational techniques to identify attacks more efficiently while mitigate the negative impact on the normal people.

In this book, we focus on providing optimization frameworks to address the above concerns in a unified manner. Specifically, with applications in diagnostic service centers and security screening respectively, we study an important aspect of service operations management: the service quality and efficiency tradeoff. In addition, we study how a manager can generate a high quality schedule in a timely manner. The remainder of the book is organized as follows.

In Chapter 2, a management problem is motivated by a type of diag-

nostic service center, in which customers have autonomy to decide whether to call the service center staffed by nurses, based on their expectation of diagnostic accuracy and the waiting time due to call congestion. Since longer service (greater service depth) typically entails higher accuracy, but also more congestion, the manager must jointly decide on the optimal service depth and the staffing level to guide and manage a diagnostic process between agents and customers. We build a model and establish its structural properties analytically. We find that the dual concerns of accuracy and congestion lead to a counterintuitive impact of capacity: Increasing capacity might increase congestion. The insight for management indicates that staffing, quality, and depth of service considerations require careful consideration in healthcare.

In Chapter 3, the second management problem is motivated by the concerns on security and congestion appearing in security screening. Security screening is widely used to identify and deter potential threats (e. g. , attackers, terrorists, smugglers, spies) among normal applicants wishing to enter an organization, location, or facility. Generally, in-depth screening reduces the risk of being attacked, but creates delays that may deter normal applicants and thus, decrease the welfare of the approver (authority, manager, screener). In this book, a model is developed to determine the optimal screening policy to maximize the reward from admitting normal applicants net of the penalty from admitting bad applicants. We provide analytical solutions for the optimal non-discriminatory screening policy and numerical illustrations for both the discriminatory and non-discriminatory policies. In addition, we discuss more complex scenarios including imperfect screening, abandonment behavior of normal applicants, and non-zero waiting costs of attackers.

In Chapter 4, we motivate a generalized scheduling problem in the context of the quality/efficiency tradeoff. A project network composed of

discretionary tasks typically exists in service professions, such as journalism, clinic, software development or financial analysis, where the quality (or value) of a task increases with the time spent on it. Since a longer task duration consumes more resources (i. e. , workers' time), the project manager must strike a balance between quality and time by scheduling tasks and setting their durations while respecting the project deadline, precedence and resource constraints. The ability to vary task durations provides a unique dimension of managerial flexibility, different from traditional purely time-based (make-span or weighted tardiness) scheduling. We formulate this problem, give a polynomial time optimal algorithm for the single capacity case, and prove the NP-completeness of the general multiple capacity case. Then we develop two hybrid solution procedures integrating linear optimization and an AI search procedure—precedence constraint posting—for the general case. Our results verify the effectiveness of these procedures and show there exists a potential synergy between objectives of maintaining temporal flexibility and maximizing quality, which implies existing techniques in building flexible schedules can be adapted to solve this new class of problems.

We present the conclusions derived from the theoretical results and experimental observations in this book in Chapter 5. Then we describe several possible extensions to enhance our models. Finally, we discuss problems which are interesting and related, but beyond the scope of this book.

Chapter 2

Diagnostic Service Centers^{①, ②}

According to statistics reported by the Centers for Disease Control and Prevention (McCaig and Burt, 2005) in the United States, during 2003, of an estimated 113.9 million emergency room visits, 13% were non-urgent. At an estimated average cost of \$300 per emergency room visit (Machlin, 2006), a cost of \$4.4 billion can be managed more efficiently by directing patients to the appropriate care center. This explains why health plans, managed care organizations, hospitals and physicians are investing in a special type of service center. Health plans and managed care organization suffer direct costs due to unnecessary visits, while hospitals and physicians suffer from the indirect inconvenience cost (staff stress, bad reputation) from the increased crowding, as well as the opportunity cost incurred by failing to provide care to other patients. This special type of service center provides high-quality decision counseling advice to patients over the phone about what the most appropriate course of action is based on their symptoms, enabling them to make more informed and thus more cost-effective treatment decisions (Sabin, 1998). We call these service centers “diagnostic service centers.” We use “diagnostic” to characterize the information (advice) provided by these service centers, which is valuable for making decisions, and is not bundled with

① This chapter is mainly based on the author's thesis and her joint work with Professor Laurens Debo, Professor Alan Scheller-Wolf and Professor Stephen Smith (Wang et al., 2010a).

② All proofs of theorems and propositions in this chapter are listed in the appendix available upon request from the author.

treatment. It is a diagnosis within nurses' capacity as professionals, *not* equivalent to medical or clinical diagnosis performed by doctors, where legal issues may be involved (Kabala, 1998). Besides the benefit of making patients feel more empowered, helping patients select appropriate services can reduce unnecessary costs incurred when patients present themselves for inappropriate treatments. Providing *help* in making these decisions is crucial; One defining feature in healthcare markets is that consumers have difficulties in identifying their needs, and thus find it hard to match their needs to the most appropriate treatment option (Arrow, 1963). As a result, a large proportion of healthcare costs are caused by unnecessary physician or emergency room visits that arise because consumers overestimate the severity of their illness (Herzog, 2003; Lynch, 2000).

Managers of such centers must strike a balance among accuracy of advice, callers' waiting time, and staffing costs by setting the appropriate staffing levels, skills sets, and service depth. Patients decide whether to use such a diagnostic based on their expectation of accuracy of advice and waiting time due to call congestion.

Diagnostic service centers have been used in many cities in the United States under different names, such as nurse line, nurse hot line, nurse triage line, telephone triage, and nurse help line. They provide a medical triage phone service, to help patients choose the appropriate care, at the appropriate place, at the appropriate time (NurseResponse, 2007). Patients may call a health coach (usually a registered nurse) 24 hours a day, 7 days a week. Health coaches provide support over the telephone to help patients interpret and act on symptoms, deciding among treatment options. For example, *Should I go to the emergency room? Should I schedule an appointment with my doctor? Should I try self-care?* A similar description of these options can be found at nurse line vendors such as