Preface

As telecommunications products and services have become an essential part of everyday life, consumers have at the same time grown intimately familiar with the concept of tiered pricing that is associated with such services. With tiered service structures, users may select from a small set of tiers that offer progressively higher levels of service with a corresponding increase in price. Tiered structures have been applied in several forms to wireless services (e.g., characterized by the amount of voice minutes, number of text messages, or the size of one's circle of friends to whom voice calls are free), Internet broadband access (e.g., the access speed or volume of monthly transferred data), and digital TV offerings (e.g., the number of channels included), among others. Service tiering is a form of market segmentation which, if applied appropriately, benefits both providers and consumers by making available services and associated price points that reflect the diversity in consumers' needs and ability to pay.

The purpose of this book is to develop a theoretical framework for reasoning about and pricing Internet tiered services, as well as a practical algorithmic toolset for network providers to develop customized menus of service offerings. We provide a comprehensive study of the design, sizing, and pricing of tiered structures for Internet services, and we illustrate their potential in simplifying the operation of complex components such as packet schedulers. The topic corresponds to a graduate-level field of study that combines the fields of Internet services, economics, and quality of service (QoS) in network resource allocation.

This book is intended for practicing engineers as well as industry and academic researchers. The potential audience includes: network designers and planners, and engineering and sales managers at Internet Service Providers involved in the selection, sizing, and pricing of tiered services, including bundled services; industry practitioners and graduate students in computer science, telecommunications, and related fields interested in Internet services and economics; and researchers who wish to explore the subject matter further. The book is also suitable as textbook for graduate-level courses in electrical engineering, computer engineering, computer science, operations research, and economics programs, that address Internet

services, Internet economics and pricing, discrete location theory, or dynamic programming optimization, as well as for industry short courses in these areas.

Book Organization

The book provides a comprehensive treatment of the problems and issues arising in providing Internet tiered services. Chapter 1 (Introduction) provides a motivation for, and definition of, tiered network services, and discusses existing business models in this context. The remainder of the book is divided into three parts, each addressing a distinct aspect of tiered services.

Part I, *Theory,* consists of Chapters 2-6. This part builds upon concepts from location theory to develop a theoretical framework for reasoning about and tackling algorithmically several important problems related to tiered network services. Chapter 2 provides an easy entry to facility location problems, and introduces a new variant, the directional p-median problem, as the fundamental problem underlying the study of tiered services. It also discusses several applications of this problem in large-scale, heterogeneous networking and computing environments.

Chapters 3 through 5 investigate the problem of optimal tier selection for services characterized by a single parameter, namely, the data rate of the access link; this problem is modeled as a directional p-median problem on the real line (i.e., one dimension). Chapter 3 considers the case where user demands are deterministic and known in advance to the network provider, and presents efficient optimal algorithms for determining the tiers to be offered so as to minimize the cost, in terms of network resources, incurred by the provider. Chapter 4 introduces the concept of "TDM emulation" that refers to a tiered-service network in which all service tiers are multiples of a basic bandwidth unit. TDM emulation is useful in several network contexts, including next generation SONET/SDH networks and traffic grooming. The tier selection problem is formulated as a constrained version of the directional p-median problem, and a suite of efficient algorithms is presented to determine jointly the basic bandwidth unit and service tiers in a near-optimal manner. Chapter 5 provides approximate yet accurate solutions to the one-dimensional directional p-median problem with stochastic demands, i.e, when only a probabilistic distribution of user demands is known.

Chapter 6 concludes the first part of the book by considering the directional p-median problem in multiple dimensions; this problem arises naturally whenever the network (1) provides a service characterized by multiple parameters, or (2) bundles several distinct services together as a single product. The tier selection problem in this case is shown to be NP-hard, and several heuristic algorithms are presented and evaluated. The main conclusions from the work presented in Part I is that a small set of appropriately selected service tiers is sufficient to approximate the resource usage of continuous-rate networks.

Part II, *Economics*, contains Chapters 7-9, and employs concepts from economic theory to formulate the problem of tier selection in a manner that takes into account

the realities of the marketplace. Chapter 7 considers a bandwidth tiered service and develops an economic model that considers three perspectives: that of users only, that of providers only, and that of both simultaneously. It also uses game-theoretic techniques to determine optimal prices for each service tier in a manner that balances the conflicting objectives of users and providers. Chapter 8 investigates tiered service as a market segmentation strategy for increasing provider profits under the assumption that consumer behavior with respect to pricing differs across the user population. Chapter 9 extends this study to service bundles, and uses tools from economics and utility theory to determine optimal service tiers when customers face budget constraints. An interesting finding from Part II is that simple tiering structures currently deployed by service providers may not work well in practice.

Part III, *Quality of Service*, illustrates the practical implications of tiered services by considering a central component of packet-switched networks, the link scheduler. Chapter 10 presents a survey of the literature on packet scheduling disciplines, and discusses the trade-offs between simplicity and ease of hardware implementation, one the one hand, and the requirement for delay and fairness guarantees, on the other hand, facing the designers of schedulers for high-speed routers. Chapter 11 then demonstrates that using the principles of service tiering it is possible to design a new scheduler, referred to as tiered service fair queueing (TSFQ), that achieves tight delay bounds and worst-case fairness with low algorithmic and implementation complexity. Experimental results from a Linux kernel implementation of the TSFQ scheduler are also presented.

Raleigh, North Carolina, USA
October 2008

George N. Rouskas
rouskas@ncsu.edu
http://rouskas.csc.ncsu.edu/