Netw Sci (2012) 1: 48–66
DOI 10.1007/s13119-011-0004-5

SURVEY ARTICLE

Internet resource pricing models, mechanisms, and methods

Huan He (✉), Ke Xu, Ying Liu

Institute of Computer Networks, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Received: 10 December 2010/Revised: 30 August 2011/Accepted: 5 November 2011
© Tsinghua University Press and Springer-Verlag Berlin Heidelberg 2011

Abstract With the fast development of video and voice network applications, CDN (content distribution networks) and P2P (peer-to-peer) content distribution technologies have gradually matured. How to effectively use Internet resources thus has attracted more and more attention. For resource pricing, a whole pricing strategy covers all related topics, including pricing models, service mechanisms and pricing methods. We first introduce three basic Internet resource pricing models through Internet cost analysis. Then, with the evolution of service types, we introduce several corresponding mechanisms which can ensure pricing implementation and resource allocation. On network resource pricing methods, we discuss utility optimization in economics, and emphasize two classes of pricing methods (including system optimization and entities' strategic optimizations). Finally, we conclude the paper and forecast future directions.

Keywords Internet, pricing, strategy, service, type, optimization, game, theory

1 Introduction

1.1 Background

Too many packets will incur network performance degradation, which is called congestion [1]. Congestion is caused by unbalanced resource and traffic distribution and thus will not be automatically eliminated with the increase of network capacity. In a packet switched network, the selfish nature of users makes this happen. As illustrated by Hardin [2], "tragedy of commons" occurs when many individuals share public resources and each holds a selfish objective, which means the loss they bring to others is larger than their own improved benefits. So, if the network is used as public goods, the overall personal excessive usage will possibly cause system performance decline and thus the congestion problem.

In recent years, with the fast development of QoS-awared video, voice and other bandwidth-consuming applications, network traffics have surged. This makes network congestion more frequent and serious. Accordingly, compared with simple priority-based QoS mechanisms [3, 4], novel content distribution technologies and multi-layer QoS mechanisms are constantly being proposed and improved upon. For the former, a new layer of network architecture, i.e., the application layer network, is added to the existing Internet, such as P2P (peer-to-peer [5]) and CDN (content distribution networks [6]). For the latter, commonly, QoS mechanisms are developed to work at multiple levels of a network, such as the transport and network layer, which are widely concerned with basic network service mechanisms, e.g., passively congestion control [7–9] and traffic engineering [10].

However, network management and performance improvement are not trivial, and they will be increasingly difficult due to the following reasons: (1) video-like traffics keep increasing as Valancius [11] shows in Fig. 1, which indicates higher QoS requirements; (2) for different CDN/P2P applications or ISPs, their selfish QoS control objectives may lead to conflicting behaviors which may even degrade network performance; (3) for multi-layer QoS mechanisms, since they often complicate network protocol design and implementation, the effectiveness is limited. Moreover, as they do not differentiate high-level application types, QoS differentiation based on service is hard to achieve. For

E-mail: heh08@mails.tsinghua.edu.cn

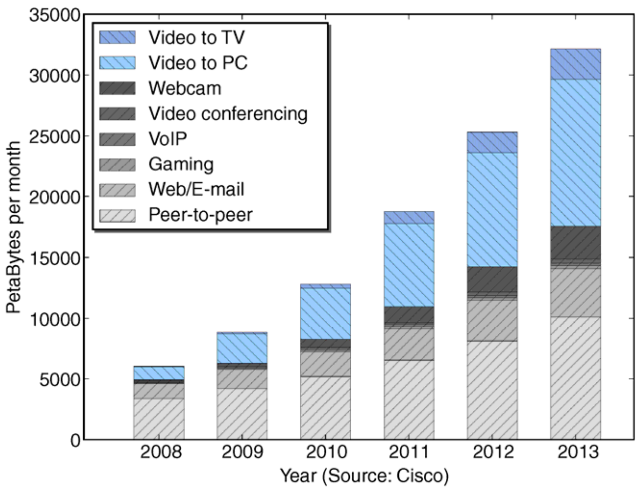

Fig. 1 Internet video content growth [11]

ISPs, a convenient way to improve QoS might be upgrading network infrastructure or increasing network capacity. However, such short-term investments usually bring high cost and fail to satisfy the fast-growing network resource requirements in the long run.

In addition, a bold attempt should be mentioned, i.e., proposing network architectures to ensure QoS. For example, IntServ [13] guarantees QoS per-flow resource reservation, and DiffServ [14] modifies IntServ architecture by adding priorities based on aggregated flow. Theoretically, they can improve network efficiency, indicating that a QoS guaranteed service era is coming. However, in addition to technical complexity, they generally achieve QoS guarantee of high priority service at the expense of the QoS of low priority service. Furthermore, as the Internet management is distributed, ISPs lack adequate enthusiasm to collaboratively improve network performance/efficiency without appropriate incentives. These largely impede the implementation of such architectures.

Then, we can conclude that when the resources are limited and some services are QoS-awared, an equally important problem with QoS improvement is how to effectively and reasonably use network resources. As for deploying new architectures, proper incentive mechanisms should be designed as a necessary support.

1.2 Resource pricing

From the above discussion, we notice that designing incentives at economical levels, to direct users to rationally use resources and to encourage ISPs in improving network performance, will be of great significance in effective network resource management and distribution [16]. Therefore, resource pricing, as an active resource management method that may affect revenue sharing among ISPs, is the key issue. Particularly, as an important auxiliary for technological progress, pricing should be suited with services.

Then, how can we realize such pricing and what are the key challenges? To answer these questions, three problems should be considered:

Q1. Basically, which factor should be charged?

Q2. How can we identify these factors in different service mechanisms?

Q3. How much should be charged?

As shown in Fig. 2, we present a complete picture of network pricing, including three aspects: basic pricing models for Q1, mechanisms to ensure pricing implementation for Q2, and methods to determine pricing levels for Q3. After we decide the pricing factors for specific services and the corresponding pricing methods, a relatively complete pricing is planned. However, obviously, the computational and technical complexities should be measured before we adopt and implement such pricing. We will briefly introduce each aspect as follows.

For Q1, we define that pricing models decide which factors to charge, or how to evaluate network operation and maintenance costs. Mason and Varian [18, 19] classified the cost as a fixed cost due to the basic service structure (such as leased lines, equipment maintenance, and human resources), marginal costs of access, network expansion costs, marginal costs of sending data packets into the congested network, and social costs caused by negative impact on other users. They believe a good price should reflect these costs. Hereby, we introduce three basic pricing models: flat pricing [18], usage pricing [20–22] and congestion pricing [18, 23–31].

Historically, when applications were simple and resources were sufficient at the beginning of the Internet, it was convenient to charge users by a flat pricing model with usage-irrelative fixed fees. However, too many packets brought by the increases of network content may have caused network resource shortages. Then, due to the lack of incentives for efficient network resource usage [32] (a lot of bandwidths wasted by non-critical applications), the overall network performance degraded. For users, the experience deteriorated and the fairness could not be guaranteed. Thus, flat pricing was no longer applicable. Then, a more effective resource pricing model “usage-based pricing” was proposed [20]. It pointed out that if the charge was usage-based, a fair and efficient use of resources would be promoted to some extent. However, with a further increase in network traffics, the aggravated congestion made the related pricing a hot research area, resulting in a relatively dynamic pricing model “congestion pricing” [18, 19] which has been studied extensively. Besides, these three pricing models could be used overlapped because they reflected different cost components.

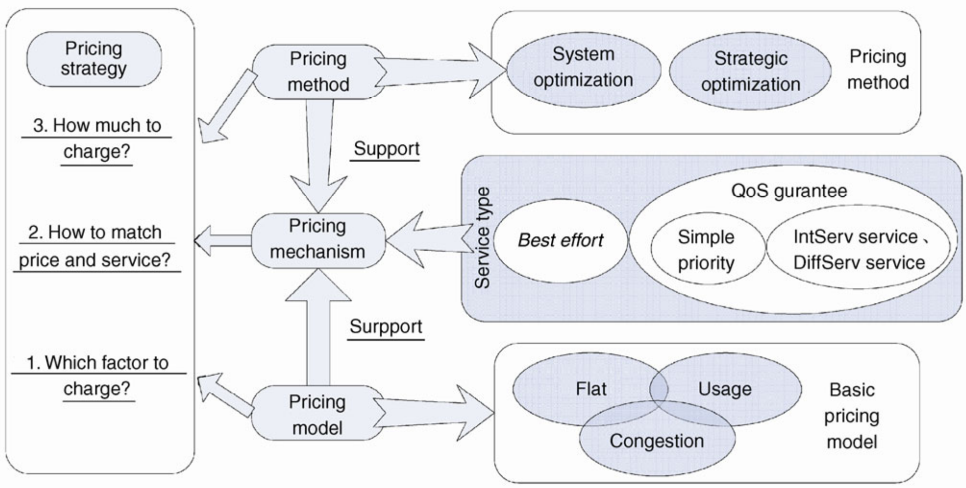

Fig. 2 The structure of pricing strategies

For Q2, we claim that pricing mechanisms mainly aim to address the matching problem between network service types and pricing models. Namely, to identify suitable pricing factors for different network service mechanisms and ensure pricing implementation with an acceptable technical complexity measure [33, 34], we simply classify the services into two types: best-effort and QoS enabled.

Specifically, in the former network, users are usually charged according to access rate or resource usage. In the latter, pricing models are adjusted to changed services. For example, Odlyzko's PMP (Paris Metro Pricing [35]) pricing aims to achieve QoS differentiation and thus enhances efficiency, so it divides the network into subnets and charges them differently. Moreover, with the increasing emphasis on QoS-awared applications and efficiency, network designers and ISPs both tend to serve different data streams with different QoS and price levels. For example, priority-based pricing was first proposed by Cocchi et al. [3, 4] to conduct service layering and corresponding pricing. Similar thoughts can be found in [36]. For QoS guaranteed network architectures (e.g., IntServ and DiffServ), the corresponding pricing mechanisms have been widely studied [37–50]. We will give more details in Section 3.

For Q3, we emphasize how to set a reasonable price level if pricing factors are identified in a specific service. In most cases, prices are results of supply-demand interactions or competitions. To achieve this goal, we will introduce various pricing methods mainly based on optimization theory and game theory. There are two major research lines:

(1) System optimization, i.e., the NUM (network utility maximization [30, 31]) framework, which is largely based on optimization theory [51];
(2) Strategic optimization of network participators, which is based on non-cooperative games [52, 53] (e.g., models in [54–58, 60]), and cooperative games [53, 61, 62] (e.g., models in [63, 65, 66]).

The rest of the paper is organized as follows (see Fig. 2). Section 2 presents three basic pricing models. Then, integrated with pricing models, we introduce pricing mechanisms based on two types of services in Section 3. Section 4 introduces pricing methods based on two classes of optimizations, including system optimization and entities' strategic optimizations in different network marketing environments. Then, we classify and compare typical pricing strategies according to pricing models, serving mechanisms and pricing methods involved. Finally, Section 6 concludes the paper and predicts future directions for pricing that is suitable for new and evolving network services.

2 Basic pricing models

In this section, we will introduce pricing factors based on cost analysis historically. Generally, there are three basic models in traditional best-effort networks and they also represent important factors in the pricing of QoS guaranteed network services.

2.1 Flat pricing

At the early stages of the Internet, users utilize a small quantity of network resources. Thus, ISPs aim to attract a large number of users and occupy the market. They generally adopt a unified price C (or flat fee [18]) to charge users based on access costs, which means that in a certain period of time, the users with the same access rate will be charged equally.

Intuitively, as the simplest pricing model, flat pricing is easy to implement and there is no need for complex statistical systems. Moreover, it can stimulate network usage since no matter how much data is transmitted, the fee is not changed. Thus, the charges can be predicted by users.

However, several drawbacks emerge as the traffic increases. As shown in Fig. 3(a)([32]), suppose the unit cost of usage is c, the charge for user is p, and the demand curve is $D(p)$, then, we can find that:

(1) Users have no incentives to limit their usage, making network resources overused: Using the flat model, the marginal usage cost for users is 0, which makes the demand changed from $D(c)$ to $D(0)$;

(2) Light users will compensate heavy users: If the flat fee C is charged based on average usage amount, then $C = c \times x_{\mathrm{f}}(\mathrm{av}) = c \times D(0)$, and all users' payments are shown as the rectangle area in Fig. 3(b). Clearly, the light-load users' payment is more than their gain, while the heavy-load users are on the contrary; and

(3) Resources are wasted to some extend: Estimated by users' practical utilities, the usage over $D(c)$ will cause $\int_0^c [D(c) - D(0)]\mathrm{d}p$ value loss, as the shade shown in Fig. 3(a).

From the above discussion, we can infer that the flat pricing model is unable to help achieve optimized resource allocation. The fewer the ISPs, the less the incentive to improve network performance. Thus, with the development of network applications and the increasing complexity of Internet marketing environment, the model will no longer work. However, as one of the referential pricing factors, access charges can be used as a basic guarantee for recovering the fixed costs.

2.2 Usage pricing

As the usage and fixed costs have been distinguished and studied separately, usage-based pricing models can be discussed. Currence et al. [21] believed that usage-based pricing can reflect actual use of network resources and is derived from traditional flat pricing. Simply speaking, usage-based pricing means the charge P is related with the amount of resource usage v, i.e., $P = p \cdot v$, where p is the unit usage price.

Usage-based pricing was studied by a lot of researchers at early stages of the Internet [20–22, 34, 67, 68]. Generally, they use a supply-demand balance model in economics to describe the interactions between users and ISPs. Edell and Varaiya [32] showed in their experiments that users are highly sensitive to pricing, and thus usage-based pricing can enhance efficiency as well as guarantee fairness among users. Moreover, experiments in [20] illustrate that dynamic usage pricing can prevent congestion and improve the average network performance. However, other problems still need to be addressed, such as the privacy issues in processing audit and statistics [21] and the charging problem caused by users' non-expected traffic (such as ads and spam).

Practically, China Education and Research Network (CERNET) uses full-rate accounting charges for international traffic [68]. Besides such direct traffic statistics, ISPs in general can use statistical sampling methods to estimate usage, such as the 95th percentile pricing. It has been used as an industry standard, and in this method, the peak flow within 5 percent of the total time (36 hours per month) is free of charge. Many ISPs, such as MCI WorldCom and Level (3) Communications, adopt such peak flow rate based charging standards [21]. Also, intelligent agents which can help users to decide resource usage based on network conditions and their willingness to pay are studied early in [69].

Recently, with the continuous development of high-bandwidth required applications and P2P content distribution technologies, the overall users' bandwidth demands have increased dramatically. Consequently, increasingly differentiated usage patterns make the fairness problem even more serious, which indicates that charging heavy-load users according to usage is more reasonable [70]. However, in terms of P2P applications' providers who encourage users to

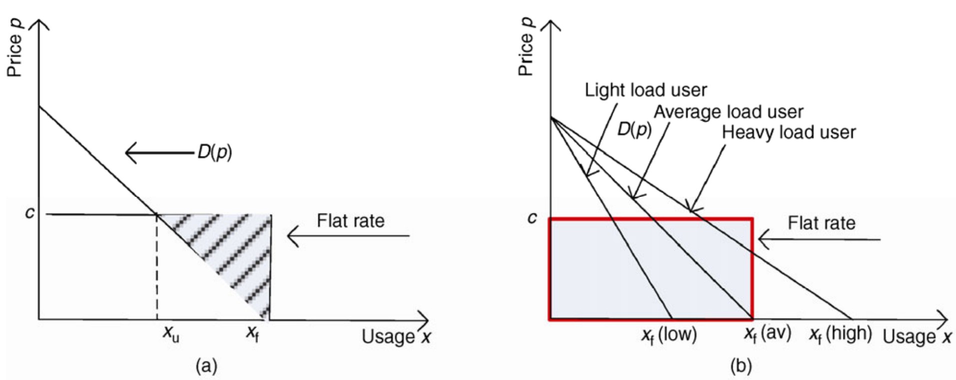

Fig. 3 (a) A customer will consume $D(p) = x_{\mathrm{u}}$ units at a unit price of p, and x_{f} under a flat-rate charge. The shaded area represents the waste [32]. (b) At a unit cost of c, the flat-rate charge is the rectangle. The small triangle is the value to the light user, and the large triangle is the value to the heavy user [32]

participate in content sharing, such charging schemes will go contrary to their goals. So, more complicated interactions between P2P application providers and ISPs should be carefully studied. For example, He et al. [71] proposed a cooperative profit distribution method to avoid such conflict. We leave out details here.

2.3 Congestion pricing

Intuitively, too many concurrent network users may lead to high system load, which is likely to affect network performance. Researchers expect to constrain this negative external effect (named social cost [18]) through pricing. In other words, when the network is busy, the pricing is used to encourage users to avoid excessive resource usage in order to relieve or eliminate congestion [18, 23–31]. The corresponding pricing is named congestion pricing.

Congestion pricing dynamically sets prices that can reflect approximate real-time network resource usage and represent current social costs. However, the measurement of such cost is not trivial. It requires detecting the user perceived value of marginal resources (like shadow price in [24, 25]) and cannot be directly calculated or measured as fixed cost or usage cost. Thus it cannot be described easily using mathematical symbols.

In general network performance optimization articles, congestion cost is described by delay in M/M/1 queuing system [56]. In Mason and Varian's smart market [18] pricing mechanism, an auction-based pricing method was proposed to measure and price such social costs. Here, the limited resources are allocated to users with high willingness to pay, so the allocation will be more efficient. However, periodic bidding and price setting makes the method more technically complex. MacKie-Mason [29] further studied the advantages of smart markets using a generalized Vickrey auction mechanism [28] to allocate scarce resources (i.e., when willingness to pay becomes a personal privacy issue, the user with the highest bidding value will get the item at the second highest bidding value). He concluded that the mechanism can promote truthful expression of users' utilities and thus help networks to attain efficiency with differentiated QoS levels. This kind of congestion pricing belongs to mechanism design (MD, [72]), which has been always studied in the incomplete information game theory area. We leave out the details here as well.

Since congestion pricing implements network-awared pricing which encourages users to shift traffics from peak time to non-peak time, congestion possibilities can be reduced. However, as mentioned above, the implementation mechanism is always complex, and its effectiveness is time-sensitive as analyzed by Ykusel and Kalyanarama [31]. They concluded that when the price interval is more than 40 times of RTT, the price can hardly affect congestion. In fact, time varying usage-based pricing can also achieve a certain level of congestion control [20], though it may not be based on the analysis of social cost.

2.4 Discussion

This section describes basic pricing models based on cost analysis in traditional best-effort networks. They are gradually proposed and thoroughly studied due to the increased usage of network resources. In a flat pricing model, the fee is generally constant in a long period of time and is used to recover the fixed cost. Usage-based fee is charged to recover usage cost. Congestion pricing is proposed to measure and charge for congestion. It is a kind of dynamic pricing where price is dynamically adjusted. Obviously, with the increasing importance of pricing in effective network resource management, pricing models will involve more factors and become more complex.

In fact, these three pricing models are not orthogonal, which means their functions can be overlapped to some extent though they reflect different pricing factors. For example, Altmann and Chu [67] proposed a hybrid pricing model that combines flat and usage pricing. In this model, users enjoy basic services at a basic flat rate, while higher bandwidth demands will be charged by usage. They show that such a pricing model can improve network performance and increase revenue for the ISP. Similarly, a "two part tariff" [19] was also proposed and it can reduce congestion to some extent.

3 Pricing mechanisms based on service type

Network service types can be divided into best-effort service and QoS mechanism related services. For different service types, pricing models should be suitable for charging [73] and mechanisms are required to ensure pricing implementation. In this section, we will use examples to introduce the matching between pricing and services and make a brief evaluation.

3.1 Best-effort service pricing

In best-effort networks, ISPs do not always implement additional QoS control which requires mechanisms along the whole transitional path. So it is reasonable that users' fees are calculated by the access network. And thus pricing is done at the network's edge, known as edge pricing [16, 73].

Typically, flat and usage pricing models are suitable for best-effort networks, but congestion pricing is not. The reasons are as follows:

(1) Flat and usage pricing are more convenient to charge users at edge network: Congestion pricing needs coordination along the whole path, which complicates the pricing; and

(2) It is also unfair to charge users according to path status: Users have no choice on routing while data packets have multiple paths to choose [73].

However, the weakness of edge pricing is that it cannot reflect network status. Thus, the effectiveness of pricing is limited. To solve this problem, several solutions are mentioned. The main idea is that ISPs can negotiate with users in an access network based on expected congestion through predicting network states [73] or estimated traffic instead of actual usage [17, 40]. Thus, the pricing is easy to implement and can prompt flexible interactions between ISPs and users (i.e., ISPs can dynamically adjust price based on network conditions and users can adjust their requirements according to their experience). Unfortunately, due to networks' distributed characteristics, although agreements exist, network-wide QoS guarantees or differentiation is hard to ensure.

However, through Paris Metro Pricing (PMP [35]) proposed by Odlyzko, we can attain prioritized services in a best-effort network. The main idea is to make users enjoy better performance at a higher possibility by paying more money. As shown in Fig. 4, the network is logically divided into channels with different transmission capacity C and corresponding price P. In principle, selecting channels at a higher price will get better service because of less competitors. And since network providers divide users into different categories through charging, differentiated services are naturally achieved to some extent. However, PMP only applies to a monopolistic network. So, if the model is extended to a complex network environment, the pricing and resource sharing should be consulted.

Similarly, Dube et al. [74] proposed a service differentiation method based on queue management. Each user chooses and joins a queue according to its price and length; while the network server implements a priority-based queue scheduling in order to achieve differentiated resource allocations. Unlike PMP, users here can estimate network congestion through queue lengths, and choose a service queue based on estimated congestion and its price.

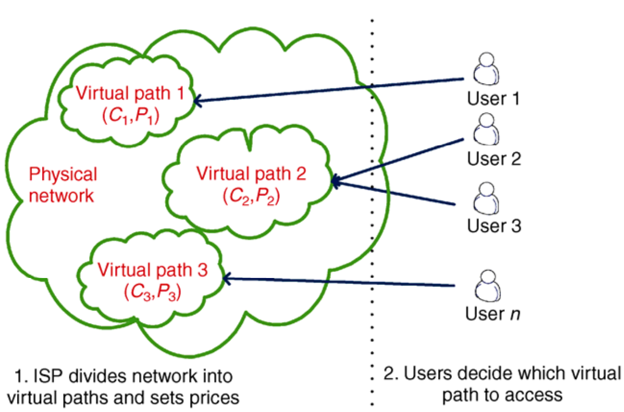

Fig. 4 The PMP pricing

3.2 QoS-guaranteed service pricing

Simple priority-based service was first introduced by Cocchi et al. [3, 4], which also revealed the relationship between QoS differentiation and resource usage efficiency. Later, the corresponding service pricing was widely studied [35, 46–49]. Then, with the development of QoS-awared applications, various in-depth studies regarding network architecture based on resource reservation [13] and flow aggregation [14] were conducted. Also, related pricing models are studied and integrated into such QoS-enabled pricing mechanisms [39, 40, 42–45].

3.2.1 Simple priority-based service pricing

To provide priority-based services, a reasonable way is to distinguish traffics by applications. Generally, packets are set to different levels of transmission priority to achieve service distinction. The simplest way is to use Type of Service (ToS) fields in IP packets to set priority levels. Such a model is more realistic and implementable though QoS may not be guaranteed. However, since packet transmission for priority-based service depends on cooperation along the whole network path, coordination among ISPs is required.

As shown in Fig. 5, services can be divided into several classes according to their QoS requirements [49]. Therefore, service with higher QoS requirements will be set a higher priority and charged at a higher price. Cocchi et al. [3, 4] showed that differentiated service and pricing can incentivize users to choose appropriate service priorities and thus prompt an efficient network resource allocation. However, since the service price is pre-set here, when idle resources exist, users will still pay more for prioritized services without the QoS guarantee.

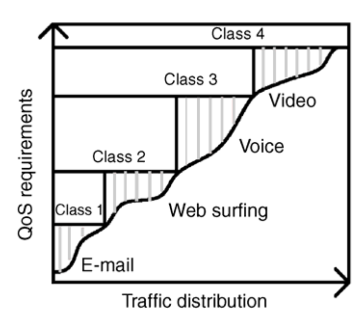

Fig. 5 Service class division based on QoS requirements [49]

Similarly, O'Donnell and Sethu [75] also suggested the priorities or service classes for data packets to be set by end users. Routers allocate them into different queues to ensure various service priorities. As to pricing implementation, the price field of a packet is filled in, which represents the payment for such transmissions. Then, when the packet reaches its destination, the price information will be copied to ACK and be returned to the sender. So the sender can determine its sending rate and dynamically select the service class based on the received price information in ACK.

Gupta et al. [46, 47] proposed a dynamic priority-based pricing mechanism and designed a real-time external price calculation method based on congestion degree in a multi-class service environment. Their simulations show that dynamic pricing can significantly improve network performance and increase revenue. Furthermore, Gupta et al. [47] studied how to set an appropriate price to prevent users from distributing traffics into non-matching service classes.

Priority-based service pricing can achieve average performance differentiation if the price and traffic are relatively stable during a long time period. However, in short term, it is likely that high-priority service will experience more packet loss, longer delays, much more serious congestion, etc. To solve this problem, a proportionally differentiated service model which provides a relatively dynamic bandwidth division scheme was studied in [49, 50]. The main idea is that, as an extension of best-effort service type, the model will not strictly set bandwidth for each service class. Instead, it will use proportional performance guarantees to achieve a predictable and controllable QoS distinction (based on well designed packet scheduling and packet discard mechanisms). Thus, the corresponding proportional pricing is more applicable to such service.

3.2.2 IntServ-based service pricing

In best-effort networks and simple priority-based service networks, QoS is not guaranteed. Accordingly, pricing usually depends on actual cost or resource usage. In contrast, this section will describe an Integrated Service (IntServ [13]) mechanism, which achieves QoS guarantee based on resource reservation. Thus, the corresponding pricing is extended from edge network to the entire resource reservation or QoS guaranteed paths.

IntServ is a single-flow based architecture that can provide an end-to-end QoS guarantee. It uses end-to-end Resource Reservation Protocol (RSVP [15]) to reserve resource for each flow. Thus, the mechanism needs all routers to process each flow's signaling messages, maintain resource reservation status, and perform flow-based classification and scheduling. Specifically, routers first convert IP packets to traffic flows and then establish/dismantle resource reservation status for each flow according to whether existing resources meet the incoming flow's QoS requirements. If so, they implement QoS routing, corresponding scheduling and other controls to ensure the required QoS based on the packets' statuses.

Karsten et al. [39] studied a pricing mechanism applicable to RSVP, as shown in Fig. 5. The main idea is to add price related information to regular RSVP messages so as to reserve resource and conciliate price. Specifically, the authors added Downstream Charging Policy Element (DCPE) in PATH messages and Upstream Charging Policy Element (UCPE) of RESV messages, where PATH and RESV are both regular RSVP messages. The mechanism works as shown in Fig. 6, and from it, we find that this pricing mechanism has much flexibility in sharing the cost between senders and receivers. Therefore, it can support pricing for many applications including one-side or two-side pay.

Similarly, Clark [40] proposed a zone-based charging or cost sharing model. In this model, a willingness to pay information is inserted into an IP packet to show whether the two sides (sender and receiver) are willing to pay for services of higher quality. Fankhauser et al. [41] proposed

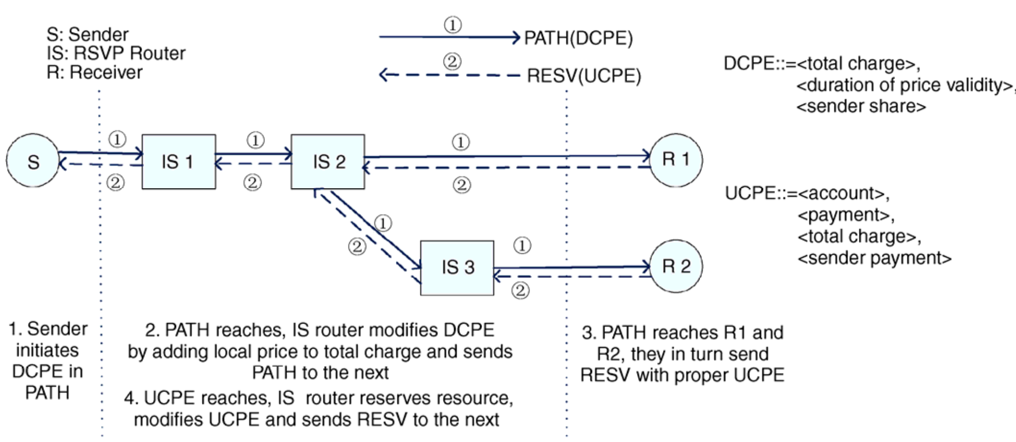

Fig. 6 Example of pricing session based on RSVP [45]

a RSVP-based accounting and charging protocol which is applicable to IntServ. They showed it can support local pricing models well through two pricing models: an auction-based pricing model (adding bid field of the RESV message) and a congestion sensitive usage-based pricing model. However, it needs to assume that the network performs static routing which will not be affected by price, and each pricing node in the network prices synchronously.

In fact, flow-based resource reservation is very complex and thus hard to achieve. It needs flow-based access control, QoS routing and related scheduling which will bring in huge system costs. Therefore, the realization of IntServ with QoS guarantee is not common and the corresponding pricing models are also under research.

3.2.3 DiffServ-based service pricing

As RSVP-based IntServ architecture has higher complexity and less scalability, Differentiated Services (DiffServ [14]) architecture is then proposed by IETF. Accordingly, the corresponding pricing is widely studied.

In DiffServ architecture, a complex flow control mechanism is realized at boundary nodes of the network and the process of inward nodes is simplified. Specifically, the boundary nodes conduct flow classification, shaping and aggregation, resulting in several flow aggregations first. Then, the aggregation information is stored in a DS (Differentiated Service) field of IP packets called Differentiated Service Code Point (DSCP). Then, the internal nodes schedule and forward IP packets based on DSCP, which represents the specified QoS requirements. As a hierarchical service structure, each DS region adopts an SLA (Service Level Agreement, i.e., a service contract between a customer and a service provider that specifies the service a customer should receive) and TCA (Traffic Conditioning Agreement) to coordinate and thus to provide cross-regional services. SLA clearly describes the supported service level and the allowed traffic volume in each service level, and TCA is used in detailed QoS negotiations.

Pricing is usually based on SLA in the DiffServ architecture. Since SLA can be a static or dynamic contract used to describe the specified QoS level on data path, the corresponding pricing can also change with SLA synchronously. In static SLA, regular consultations are needed; while in dynamic SLA, users need signaling protocol (e.g., RSVP) to help request service dynamically. Transformation is needed to match service requirements with DSCP value (no matter by user or edge router). Then, accordingly, the price for differentiated service depends on SLA and actual network resource usage. Fankhauser and Plattner [42] proposed an implementation profile to describe resource transactions in networks, which is based on the bandwidth broker to act as an SLA trader or negotiator. The essence is that through negotiation between bandwidth brokers of each adjacent ISP, an ISP can provide its neighbors with its own network resources as well as the resources it purchased from other adjacent ISPs. Therefore, the Internet-wide communication can be achieved. For example, in a core network, as shown in Fig. 7, there are six DS domains: A, B, C, D, E and F. Each DS domain represents an ISP. For network A, B may offer access to network E if it has bought access from network C or D to destination E. In an access network, as shown in Fig. 7, if user G (in network A) and network A arrive at an SLA that G will communicate with user H in network F, then an end-to-end service can be attained by building up bilateral agreements step-by-step in the form of SLAs between adjacent networks.

Furthermore, Semret et al. [43] established a double-layer DiffServ-based market model which considered users, bandwidth brokers and bandwidth sellers in one market. Each service class has its own bandwidth broker that belongs to the bandwidth seller. They concluded that competition among bandwidth brokers would lead physical bandwidths to an effectively division for various classes of service. Users adopt SLAs to negotiate services and prices with bandwidth brokers. Driven by a dynamic market, bandwidth division among various service classes will finally be stable.

Similarly, Wang and Schulzrinne proposed a framework named Resource Negotiation and Pricing (RNAP) [44]. They pointed out that pricing for reserved resources should be conducted on two different levels. In an edge network, users and ISPs negotiate based on a single flow; while in the core network, users' requests with the same service level and consultation interval are aggregated to process together. Finally, network resources are allocated based on the single flow in the edge network. In [45], Wang and Schulzrinne built an optimization model to study pricing and the corresponding implementation which introduced access control to aid resource allocation. They analyzed the resulting resource utilization in a differentiated service network. The authors concluded that congestion-sensitive pricing combined with user-controllable traffic rates can not only achieve congestion control to a large extent but also

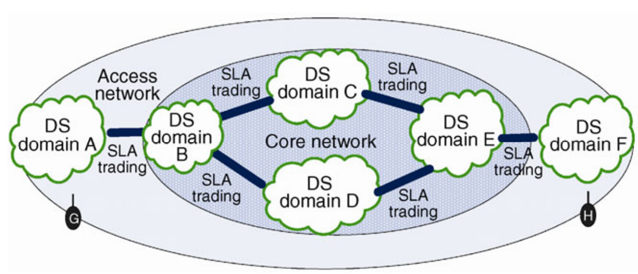

Fig. 7 Example of ISP networks at access and core levels. [42]

guarantee QoS of different service classes. Since all routers participate in congestion pricing along a transmission path, their work is more complex than the distributed pricing proposed by Yukesl [38].

In [76], the authors proposed a pricing mechanism that distinguishes core network and edge network. They claimed to charge users on the access side with a Time of Day (TOD) price which can dynamically reflect the congestion degree in core networks. For core networks (as shown in Fig. 7), dynamic pricing based on congestion for differentiated services is studied, where adaptable prices are published as signs of core network congestion status. Its advantages are as follows: (1) Since access control can be conducted in the user end system or edge network, it reduces network control information transmission and simplifies the core network processes; (2) On the other hand, as this pricing is based on DiffServ and concerns economic objectives and resource usage efficiency, it is easy to achieve a certain level of economic efficiency when providing QoS differentiated services. So, it is a flexible, scalable and efficient pricing mechanism in DiffServ architecture.

3.3 Discussion

Based on two types of network services considering QoS or not, we introduce two kinds of pricing mechanisms in this section.

For best-effort service, it is believed that if edge pricing uses expected congestion information, it can achieve a certain degree of congestion control. Also, one can distinguish access bandwidths to provide some kinds of prioritized services. But both of them cannot assure the usage efficiency of resources and guarantee QoS.

It is more complex for QoS-based pricing, because QoS is differentiated by packet processing according to service classification or resource reservation, which often needs support from devices or networks on the entire transmission path. The corresponding pricing process can be more difficult with higher complexity, especially for IntServ pricing where QoS is guaranteed based on per-flow resource reservation. However, for DiffServ, QoS is guaranteed based on aggregated flows, so it improves the efficiency at a lower complexity compared with IntServ. In fact, combining IntServ (in edge network) with DiffServ (in core network) to provide differentiated services can enjoy the benefits of the both.

4 Pricing methods

In microeconomics, price level is calculated based on related pricing theories (profit/social welfare [77] maximization) and depends on market environments or structures (such as monopolistic or competitive networks [46]). In the network research area, besides considering the market, resource pricing is also affected by network service mechanisms, and is generally settled through interactions among various entities who optimize their utilities.

In this section, we will introduce two main network pricing methods which determine appropriate price levels: (1) System optimization models which are mainly based on a network utility maximization (NUM [25, 26]) framework; and (2) Strategic optimization models which describe strategic behaviors of participators.

4.1 Pricing based on NUM

From an economic point of view, an efficient market means that total social welfare (i.e., the sum of service providers' surplus and users' surplus) is maximized [77]. Under different market environments, different conclusions can be drawn. Wemainly introduce system utility (social surplus/welfare) optimization oriented pricing methods for a single network (affects from other providers are avoided) based on optimization theory. The system consists of users with different utility functions and a network with resource constraints [23]. In fact, this research line has had a tremendous influence on communication networks. It promotes an in-depth understanding of network architecture and guides protocol design for more efficient network resource usage.

4.1.1 System model for elastic traffics

Kelly [25] proposed the concept of Network Utility Maximization (NUM) which is the initial work of Internet system optimization. In his work, the main object is to find the price that can make the total resource demand and supply in equilibrium. According to market pricing theory in [46], if a system is in equilibrium, the system utility or social surplus will be maximized. NUM framework can be described by three optimization problems. The system optimization is shown as follows:

$$\begin{array}{ll} \text{A: SYSTEM}[U,H,A,C]: & \\ \text{maxmize} & \sum_s U_s(x_s) \\ \text{subject to} & Hy = x, Ay \leqslant C \\ \text{over} & x, y \geqslant 0 \end{array} \tag{1}$$

where x_s denotes the traffic rate, U_s denotes the value or utility of the traffic to user s. The service provider's cost is ignored. Then, the constraints are: (1) $Hy = x$, where $H_{s\times r}$ denotes the source-destination pair $i \in \{1,2,\dots,s\}$ served by path $j \in \{1,2,\dots,r\}$, and vector $y = \{y_1, y_2, \dots, y_r\}^{\mathrm{T}}$ denotes the resources distributed to all source-destination pairs on each feasible path. This constraint means the whole distributed resources are equal to x_s for any user; (2) $Ay \leqslant C$, where A is a 0-1 matrix telling whether the distributed resource is on the link, and the constraint means the sum of all distributed resources will be no more than link capacity C; and (3) $x, y \geqslant 0$.

Since user utility is unknown to the system, it is difficult to solve (A). In NUM, Kelly shows that the solutions of (A) are the same as those of two sub-optimization problems: user-side optimization (B) and network-side optimization (C).

$$\begin{aligned} &\mathrm{B: USER}_s[U_s;\lambda_s]: \\ &\text{maximize} \quad U_s(m_s/\lambda_s) - m_s \\ &\text{over} \quad m_s \geqslant 0 \end{aligned} \tag{2}$$

where λ_s denotes the price of per unit traffic charged to user s. Here, user s optimizes his surplus $U_s(m_s/\lambda_s) - m_s$ by deciding how much to pay m_s (which can be indirectly inferred by $x_s = m_s/\lambda_s$). For network, it allocates network bandwidth to different flows according to users' feedbacks and some fairness standards shown as follows:

$$\begin{aligned} &\mathrm{C: NETWORK}[H, A, C; m]: \\ &\text{maxmize} \quad \sum_s m_s \log x_s \\ &\text{subject to} \quad Hy = x, Ay \leqslant C \\ &\text{over} \quad x, y \geqslant 0 \end{aligned} \tag{3}$$

where H, A and C denote the network status with the same meaning in Eq. (1). Then, given $(m_1, m_2, \ldots, m_s)$, it tries to distribute bandwidth by solving (C) which seems based on weighted proportional fairness. Kelly pointed out that if $\forall s, U_s(\cdot)$ is concave, then this convex optimization problem has a unique optimal solution $x^* = (x_1^*, x_2^*, \ldots, x_s^*)$. As $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_s^*)$, and $m^* = (m_1^*, m_2^*, \ldots, m_s^*)$, $m_s^* = \lambda_s^* x_s^*$ holds for every $s \in S$, the three optimization problems are all solved with consistent solutions. The vector x^* is the unique optimal allocating rate and λ^* is the current optimal resource price vector.

System optimization problem (A) can also be decomposed into other types of sub-optimal problems. As its essence will not change, we just skip it here. Kelly [25] further discussed the stability of the above mentioned rate allocation algorithm when random disturbance and delay were added into the system. For concrete solutions, since Kelly mainly modeled the elastic system, where users' utilities are all concave functions (it is reasonable when modeling traditional data services, such as file transfer, which is not very sensitive to delay), optimal solutions can be obtained based on convex optimization theory. Besides, the authors in [26] discussed a method that uses underlying buffer management to implement end-to-end proportional resource allocation, which supports Kelly's work.

Unlike centralized resource allocation method mentioned above, Ozdaglar and Srikant [93] pointed out that for distributed resource allocation, achieving system goals requires that: (1) End users adjust their rates according to congestion feedback sent from the network (indicated by prices); (2) Network routers calculate price which can reflect congestion status of each link starting from them; (3) Network returns congestion information (price) to users. For example, based on the fair end-to-end congestion control mechanism proposed by Mo and Walrand [9], La and Anantharam [78] proposed a distributed algorithm where users can determine their rate adjustments according to their perceived network status. In their work, each user pays for queuing delay/packet loss rate caused on others by its own packets. The authors proved that the algorithm is convergent and showed that it can be used to deduce an optimal solution. However, in engineering, how to control rate based on the price is still not well resolved.

4.1.2 *System model with inelastic traffics*

We have discussed the system model for elastic flows, where users' utilities are always described as concave functions. However, in fact, such willingness will vary with different types of applications. For example, for video and voice applications, if the transmission rate is less than a certain value, the user's experience will decline sharply (as shown in Fig. 8). This indicates that S-type utility function should be used to model user's utility. Thus, the convex optimization framework of NUM will no longer work. The resulting system can be seen as a hybrid service system, which also includes inelastic flows. Therefore, the pricing and resource allocation problem becomes a difficult non-convex optimization problem [79–82].

Jang-Won et al. [79, 80] first showed that in a real network environment (i.e., hybrid service systems), if the flows are all modeled by concave utility functions, under the NUM framework, the resulting rate allocation will probably cause

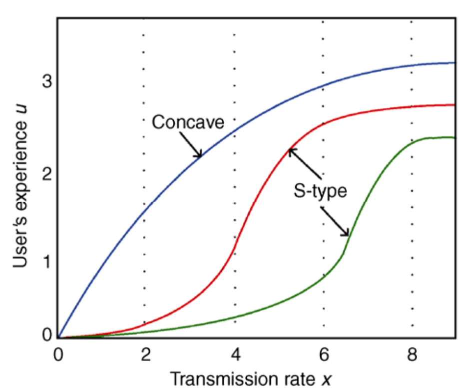

Fig. 8 Hybrid service system with various utility types

network congestion and high jitter. Then, to achieve the optimal system resource usage when heterogeneous flows coexist, they designed an incentive mechanism to inspire users' transmission cancellations when needed. Such user behavior is called "self-regulate", which is similar to the end system access control. Mathematically, as the problem is non-convex and the duality gap may exist, the solution may not converge to the optimal one. Thus, an asymptotical optimal resource allocation algorithm is further designed by them.

Unlike the above approximate optimal solution, Chiang et al. [82] and Hande et al. [83, 84] studied rate allocation optimization framework for inelastic flows, and presented the sufficient and necessary conditions for the convergence to the global optimum of the proposed distributed rate allocation algorithm. In contrast to the work by Jang-Won et al. [79], Chiang et al. [82] generalized user utilities for different types of time-sensitive flows. They modeled them using non-convex optimization tools and proposed a heuristic access control algorithm and a rate allocation algorithm. Similarly, considering the real-time flows, Hande et al. [83, 84] introduced a price-based distributed access control method and proposed a fair resource distribution method when various types of flows coexist. It emphasizes QoS-guarantee for elastic flows and is realized by a proposed heuristic algorithm.

In fact, the NUM framework has also been applied to edge pricing model because some researchers consider that access resource is most scarce and should be the focus of study. For example, based on NUM, Hande et al. [85] studied the edge pricing in a monopoly market where an ISP aims to maximize its revenue, and the user utility is modeled by standard α-fairness based on different demand elasticities, namely:

$$u(x) = \begin{cases} (1-\alpha)^{-1} x^{1-\alpha} & 0 \leqslant \alpha < 1 \\ \log(x) & \alpha = 1 \end{cases} \tag{4}$$

where α indicates how elastic the user perceives rate x. Unlike Kelly's work, Hande et al. [85] emphasized that in edge network, pricing structure can be a linear combination of time-related flat fee and usage fee. They analyzed each part's effect on the ISP's revenue.

Currently, the sender and receiver (supplier and demander) may have different utilities to traffic between them. So, ISPs need to set a supply-demand balanced price to maximize their revenue. Hande et al. [86] extended the NUM framework by adding content providers (CPs) to the system model. They concluded that no matter under which network marketing environment (competition or monopoly), if CPs are charged to compensate users, the overall system revenue and the utility of CPs will surely increase. They also discussed network neutrality issue (NN [87], simply speaking, ISPs should not charge CPs differently based on content type).

4.2 Pricing based on game theory

Within a single ISP network, system equilibrium based on supply-demand relationship is achieved through pricing where the ISP and users indirectly interact with each other. However, in real networks, there are three types of relationships: ISP-ISP, ISP-users, and user-user. Most of them are modeled by considering their direct interactive effects based on game theory, which studies how individual decisions are made when others' actions are considered. Based on whether a binding agreement can be formed, games are divided into non-cooperative games [52, 88] and cooperative games [61–62].

4.2.1 Non-cooperative game model

Considering non-cooperative games in network resource pricing and allocation, three levels of such interactions can be identified: (1) Competition among Multi-ISP in network market. As users purchase services from the most attractive ISP, when an ISP decides price, it should consider the other ISPs' behaviors as well. (2) Leader-follower game between ISP and users. If ISP considers the users' reflections directly (instead of resting on resulting demand as shown in Section 4.1), such interaction can also be regarded as a game. (3) Resource competition among users.

Research on Multi-ISP interaction faces great challenges. Besides similarities and differences among ISPs, their impact on underlay user behaviors should also be considered [94]. Therefore, mature research results remain in shortage today. In this section, we will mainly introduce non-cooperative game models for (2) and (3). Two basic theoretical models frequently used here are N-person non-cooperative game model and leader-follower game model. The former mainly considers static game equilibrium, and the latter emphasizes dynamic processes of a game.

It is reasonable to study the above mentioned relationships in a single ISP network where interferences from other ISPs can be largely avoided. Then, for the modeling of relationship (2) in a monopoly network market, a leader-follower game model (such as Stackelberg [54–58, 89]) is always used. According to how much users' utility information is known by the ISP, such work can be divided into two kinds: pricing with complete or incomplete information. For the modeling of relationship (3), an N-person non-cooperative game is always used. Here, each one's behavior affects others' utilities, which is similar to the externality mentioned in the foregoing discussion on congestion pricing.

Generally, in a leader-follower network resource pricing model, the leader (ISP) sets price strategically, and the followers (users) act as price takers, who decide how much resource to buy mostly based on the given price. The point here is that when the leader decides a price, it sets one that can maximize its revenue based on the predicted users' reflections. In the N-person non-cooperative game, the stable state (i.e., Nash Equilibrium [52]) where none of participates wants to deviate from its behavior when the others' strategies are known, is the major concern. An instance that combines the two models is presented by Basar and Srikant [55–58].

Specifically, in [56], the authors used non-cooperative game models to study pricing issues in a single-link network. They built two layers of games: a non-cooperative game related to resource competition among users and a Stackelberg game where an ISP maximizes its revenue within constrained resource based on predicting users' reflections. In the first layer model, each user i maximizes the following Eq. (5) to decide its traffic rate x_i :

$$F_i(x_i, x_{-i}; p) = w_i \log(1 + x_i) - \frac{1}{nc - \sum_j x_j} - p x_i \tag{5}$$

where nc is link capacity, $w_i \log(1 + x_i)$ is user utility function, $\frac{1}{nc - \sum_j x_j}$ represents congestion cost (i.e., queuing delay computed using M/M/1 queuing model), and p is the unit price charged by ISP. Then, they proved that such non-cooperative game has a NE, i.e., for any user i, the solution x_i^* holds:

$$\underset{0 \leqslant x \leqslant nc - x_{-i}^*}{\text{maxmize}} F_i(x_i, x_{-i}^*; p) = F_i(x_i^*, x_{-i}^*; p) \tag{6}$$

In the second layer game, authors hypothesized that the ISP aims to maximize its benefit by solving Eq. (7), and thus obtain the optimal unit resource price p.

$$\underset{p \geqslant 0}{\text{maxmize}} L(p; \overline{x}^*(p)), L(p; \overline{x}) := p \cdot \overline{x} \tag{7}$$

where $\overline{x}^*(p) := \sum_i x_j^*(p)$ represents the sum of all individuals' rates at the above mentioned NE. Then, according to Eq. (5), since adding up all utility functions of users would not change the NE point, they derived an equivalent optimization problem for users in Eq. (8):

$$F(x_1, \ldots, x_n; p) = \sum_{j=1}^{n} w_j \log(1 + x_j) - \frac{1}{nc - \overline{x}} - p\overline{x} \tag{8}$$

where all utilities are added together. Then, by solving this convex optimization problem, a unique optimal solution $\overline{x}^*(p)$ can be obtained (notice that the solution is a function on price p). Finally, a single-variable optimization problem can be obtained by substituting the above solution to Eq. (7). Solving it directly can obtain the optimal price p^*. They also considered different link bandwidths, and analyzed how the price, revenue and user utility relate with each other. They claimed that if the ISP expanded bandwidth in proportion to the number of users, the revenue would increase accordingly. Under certain circumstances, the solution will be consistent with Kelly's system optimal solution based on NUM model. The authors gave an extended discussion in the case of multi-link afterwards [55].

Similar to the above mentioned non-cooperative game framework, Shen and Basar [58] extended the model to study non-linear optimal pricing in the cases of complete and incomplete information of users' utilities. They concluded that in the former (users' utilities are known by ISP) case, non-linear price can increase ISP's revenue by 38% compared with the revenue gained by linear price; while if users' utilities are unknown (incomplete information), the loss of benefit will be 25%–40%. Li, et al. [90] also considered optimal pricing in a monopoly market with incomplete information. But they did not directly model users' non-cooperative behaviors.

However, when an ISP prices users, in addition to considering the users' response strategies, the market environment is also taken into account. For example, in a multi-ISP market, ISPs compete for users, and their prices are affected by each other. Thus, the model will become more complex. Acemoglu and Ozdaglar [60] claimed that unlike the monopoly case where system efficiency can be improved and the social optimal is achieved at the equilibrium, in a multi-ISP competition game [77], pure strategy NE may not exist (depending on cost function). Then, unlike the conclusions drawn from economics, an increasingly competitive market will reduce system efficiency. Besides, the upper and lower bounds of possible loss are also discussed in [60].

4.2.2 Cooperative game model

Historically, the well-studied cooperation game models in network resource pricing are Nash Bargaining Game [61] and Shapley value [62] models. These two models both belong to the axiomatic method, and thus their solutions satisfy certain properties. Especially, the former emphasizes Pareto optimal property and a certain level of fairness; the latter has well-formulated marginal contribution concept and the corresponding calculation methods. In recent years, as a new trend, such cooperative game models are studied and gradually applied to the modeling of network resource pricing [63–66, 71, 95].

1) The NBS model

Generally, NBS satisfies all the following four axioms [53, 61, 64]: (1) Invariant to equivalent utility representations; (2) Pareto optimality; (3) Independence of irrelevant alternatives; and (4) Symmetry. Therefore, it is usually applied to incentivize efficient and fair cooperation.

In [63], Cao et al. assumed that all network users have the same behavior characteristics and preferences. Thus, they simplified pricing problem as a game between a single user and an ISP where each of them conducts its own optimization. Then, they studied the results of a leader-follower game and a NBS model, and through theoretical analysis, they concluded that Nash bargaining can make the system operate at Pareto efficient point (one cannot increase its utility without reducing others' utilities) with guaranteed fairness compared with the results in the leader-follower game.

Furthermore, in [64], based on Nash bargaining game, the authors studied distributed network resource pricing and allocation within a network with multiple heterogeneous users. It is a more realistic and implementable example, so we briefly introduce it as follows:

First, the ISP faces a centralized fair resource allocation problem which is formulated in accordance with the concept of Nash bargaining. It is shown as the following constrained convex optimization problem:

$$\begin{aligned} \text{maximize} \ & \prod_{i=1}^{N}(x_i - MR_i) \\ \text{subject to} \ & x_i \geqslant MR_i \\ & x_i \leqslant PR_i \\ & (Ax)_l \leqslant (C)_l \end{aligned} \tag{9}$$

where x_i is resource (rate) assigned to user i, MR_i and PR_i are the minimum and peak rate requirements of user i, respectively. Based on optimization theory, it is easy to know that there is a unique optimal solution. However, such a central solution always brings in a lot of network communication burdens. Therefore, the authors proposed a distributed model where each user optimizes its utility with an added penalty $\alpha_i x_i$, and the aggregated rate is expected to ensure that the system can operate at Pareto optimal point. Thus, for each user, it optimizes Eq. (10) for rate selection:

$$\begin{aligned} \underset{x_i}{\text{maximize}} \quad & \ln(x_i - MR_i) - \alpha_i x_i \\ \text{subject to} \quad & x_i \geqslant MR_i \\ & x_i \leqslant PR_i \end{aligned} \tag{10}$$

Similar to the leader-follower game in Section 4.2.1, the network here needs to solve the rate allocation problem which can maximize its revenue. Besides, the revenue is calculated by the sum of penalties, as shown in Eq. (11). The constraint conditions are the same as those in Eq. (9).

$$\text{maximize} \sum_{i=1}^{N} \alpha_i x_i \tag{11}$$

The authors designed and implemented an asynchronous distributed algorithm with the corresponding information exchange method, and showed that the solutions of Eq. (11) by network and Eq. (10) by users are equal to Nash bargaining solutions of the centralized problem in Eq. (9). Obviously, such a distributed method can maximize users' utilities as well as the network's revenue, which is similar to the results of NUM-based model. However, the key difference is that their system objectives are different: one is to maximize social welfare and the other is to fairly distribute resources.

In [71], similar to the idea in [63], the authors studied how much improvement cooperation can make compared with non-cooperation in network pricing games. However, they mainly studied the interactions among three kinds of players: ISP, CP (providing P2P service) and the user. A multi-leaders-follower game (where ISP and CP act as leaders, and user acts as follower) was built to act as a referential model. After having built and solved a cooperation model between ISP and CP based on NBS, they computed the utility improvement of each player and concluded that such a cooperative method not only guarantees fair profit distribution, but also helps to improve the economic efficiency of the overall network system.

2) The Shapley value model

Recently, Shapley value is increasingly applied in network field, mostly in cost sharing or revenue distribution among multiple cooperative network participators [62] (e.g., cooperative ISPs in [65, 66]). Unlike in non-cooperative games where each ISP charges users directly to gain revenue, Shapley value emphasizes revenue distribution based on weighted marginal contribution of each entity in a group. As an axiomatic method [62, 65, 66], basically, it satisfies the following properties: (1) efficiency; (2) symmetry; (3) fairness; and (4) dummy.

Proposed by L. S. Shapley in 1953, the Shapley value φ provides a unique payoff allocation satisfying some fairness criteria. It is defined by

$$\varphi_i = \frac{1}{N!}\sum_{\pi \in \Pi} \Delta_i(v, S(\pi, i)), \quad \forall i \in N \tag{12}$$

where Π is the set of all $N!$ permutations of N, and $S(\pi,i)$ is the set of players preceding i in the permutation π. Thus, the Shapley value of each player can be explained as the expected marginal contribution $\Delta_i(v,S(\pi,i))$, where S is the set of players preceding i in a uniformly distributed random permutation.

There emerge several studies on applying this theory to real network pricing or profit-sharing. For example, in [65] and [66], through computing marginal contribution of each ISP, the authors presented a revenue distribution method based on Shapley value as shown in Eq. (12). However, generally, high computational complexity is its obvious drawback (i.e., N participants needs 2^N scale of computations). Besides, it requires a centralized allocation process which will also make it less scalable. Besides, in [95], Misra et al. proposed to fairly share profits among all involved parties based on Shapley value to incentivize peer-assisted services.

4.3 Discussion

We classify and summarize typical pricing methods of network resources based on two main research lines. The key points are as follows:

(1) System optimization models which are mainly based on the NUM framework. Considering network traffic characteristics, they can be divided into: i) Optimization models for elastic flow system; ii) Optimization models for a hybrid system where inelastic and elastic flows coexist.

Generally, their goals are to find the optimal price and rate allocation with balanced supply and demand where the maximal (in elastic flow system) or approximate maximal (in most hybrid system) system efficiency is achieved. Therefore, price-based access control is mainly introduced in the hybrid system to assist resource distribution. It generally includes two methods: users' self-regulation [80] and access control from the network [83]. However, since access control policies of each link may be different on a network transmission path and there lack well-designed distributed decision-making mechanisms, system convergence cannot be ensured. Elastic flow protection in hybrid systems was proposed by Hande et al. [84], who believe that elastic traffics are less competitive than inelastic flows.

(2) Strategic optimization models which are based on two major branches of game theory: non-cooperative game and cooperative game. However, non-cooperative game model mainly discusses NE and its characteristics, including efficiency, uniqueness and stability; while cooperative game model emphasizes fairness, Pareto efficiency and other expected properties. For the latter, third-part supervision is always required to enforce cooperation.

Actually, if users can anticipate the effects of their actions, the system optimization will become a game [91, 92]. This briefly describes the relationship between system optimization model and non-cooperative game model we introduced here. In the former, the equilibrium is achieved by indirect interactions between the network and users based on price. In this process, ISP dynamically controls the system through pricing mechanism to help it reach an optimal equilibrium. In the latter, analyzing strategic behaviors of all participants based on non-cooperative game theory can help determine whether the system has NE, and evaluate its efficiency. As stated in [91] and [92], such game leads to an aggregate utility that is no less than 3/4 of the maximum system utility (for more detail, please refer to [92]).

5 Classification and comparison of pricing strategies

In this section, based on pricing models, service mechanisms and price level setting methods, we conduct classification and comparison of the introduced typical pricing strategies shown in Table 1. In order to describe the pricing for QoS guaranteed service, we add QoS contract in the pricing model, which represents the agreements between ISPs and users on service and price.

In Table 1, it should be noted that early pricing models lack theoretical basis, and most of them are based on experiments, so they cannot cover a complete decomposition. Some articles focus on studying pricing methods without QoS differentiation, so we generally assume they are applicable for best-effort networks. In addition, the QoS guaranteed types of services refer to what we have described in Section 3.2. For a pricing model, if both usage and access are chosen, it is a composited one.

Obviously, pricing for different service types inherently have different technical complexities. Generally, for best-effort network, pricing is always done at network edge, and incurs less overhead cost; while for QoS guaranteed services, since pricing relates with QoS along the whole serving path, it involves higher audition and accounting cost to achieve higher network efficiency or better performance.

We show several examples to reveal how different ingredients can be combined for building a whole pricing strategy. In [70], Wang et al. studied pricing in best-effort network using a composited pricing model based on a two-player non-cooperative game; while in [59], Altman et al. studied pricing in differentiated services and its impact on system equilibrium based on non-cooperative game theory. Especially, for pricing that is based on Shapley value in cooperative game model, marginal contribution is the only measurement for payoff, so we leave out such work in Table 1.

From Table 1, we can conclude that as the service and market environment become more complex, the corresponding pricing factors and pricing methods will become even more complicated. Also, we present the evolution process of pricing strategies as Fig. 9 shows. It is a manifest trend that when multi-ISP and multi-CP are involved, the game model will be a more suitable and attractive choice.

Table 1 Classification of pricing strategies

Pricing model				Service type				Pricing method			
					QoS-guarantee				Game model		
Access	Usage	Congestion	QoS contract	Best effort	Priority	IntServ	DiffServ	System model	Non-co	Co	Example
	√			√							[20, 21]
	√			√				√			[22, 69]
√	√			√							[67]
√	√			√					√		[70]
√	√			√				√			[19]
	√	√						√			[24, 25][75]
		√							√		[29](MD)
		√		√				√			[30]
			√	√							[16, 17, 73]
√				√				√			[35]
	√	√			√			√			[74]
	√				√			√			[3, 4] [75]
		√			√			√			[46, 47]
			√			√					[39, 40]
	√	√				√			√		[41](MD)
	√	√					√		√		[43](MD), [59]
		√	√				√	√			[45]
		√				√	√	√			[76]
		√		√				√			[78]
	√	√		√				√			[79–84, 93]
√	√			√				√			[85]
	√			√				√			[86,90]
	√			√					√		[55–58, 60]
	√	√		√						√	[63, 64]
	√			√						√	[65, 66]

NOTE: The symbol √ in each row represents a feature hold by the pricing strategy example in the last column, and the symbol (MD) means mechanism design.

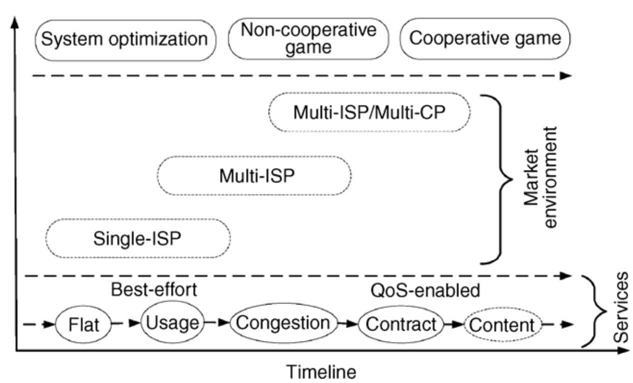

Fig. 9 The evolution of pricing strategies

6 Conclusions and future work

In recent years, with the continuous development of high-bandwidth applications, content distribution technologies (such as CDN and P2P) are increasingly mature, and network traffic surges. Thus, network service quality has drawn more and more attention. Resource management and congestion control tend to have high technical difficulties in engineering, which make network performance guarantee and maintenance even harder. However, network resource pricing can alleviate or even resolve this problem by actively affecting resource demand and usage, and thus has other important research values in addition to economic goals. Besides, as QoS-guaranteed services are getting more mature, pricing also acts as a more important auxiliary to incentivize technological progress. Thus, it is equally important to study pricing strategies that are applicable to continuous renewal services.

As shown in Fig. 2, we survey pricing issues from three different perspectives. Firstly, we introduce three basic pricing models (including flat pricing, usage pricing and congestion pricing) and conclude that with the development of network applications, such model will become more complicated. Secondly, we introduce the concept of pricing mechanism, which combines pricing model with service. The mechanism aims to ensure pricing implementation under certain service, such as transferring price information in a DiffServ network. We also notice that resource management for QoS differentiated network mainly uses price-based access control. Thirdly, in setting the price, we highlight two methods here: system optimization based on the NUM framework and strategic optimization based on game theory. We conclude that non-cooperative game models are often resting on relevant optimization theories to prove the existence of Nash equilibrium and study its properties. In addition, due to incomplete information in such game, there is still a long way to realize its practical application. Also, for cooperative game models, although efficiency and fairness are quite satisfactory, third-part supervision is always necessary.

To sum up, with the fast development of applications, service types and corresponding theories, pricing related issues are constantly updated and studied. However, whichever pricing strategy we adopt, the basic pricing models and methods can hardly change. For example, when flat pricing brings tolerable system efficiency loss, it should be revalued due to its simplicity [89]. Through extensive study on network resource pricing strategies and deep analysis on the status quo, we can draw the following conclusions:

(1) Network resource/service pricing can be used as an effective tool to prompt technical progress, support QoS improvement, and enhance network efficiency economically.
(2) Pricing model for QoS differentiation is still a hot research point, which also needs the support from the corresponding service mechanisms.
(3) Pricing is expected to be scalable and easy to implement. Besides mature theoretical models, well-designed mechanisms should also be implemented to help achieve pricing goals (such as maximizing social welfare or economic efficiency).
(4) ISPs' pricing indirectly affects service quality and the traffics of network users. Then, fair and implementable cooperation mechanism with win-win results among ISPs will be an increasingly hot topic for future research (e.g., [65, 66, 71]).

Furthermore, the models discussed above are mostly for unilateral market whose network services include content provisions. But, if content providers and users are separately considered, then under such a bilateral network market with multi-ISP and multi-CP, pricing will involve more complex interactions (as introduced in [71, 86]). Also, the network neutrality concept [87] has recently been proposed, which causes more debates on whether the content should be charged differently. We can infer that content-based pricing will likely be further discussed as a part of pricing models in the near future.

Acknowledgements

The authors are grateful to the anonymous reviewers for their constructive and valuable comments that helped in improving this paper. This work was supported in part by NSFC Projects (60970104 and 61170292), 973 Project of China (2009CB320501), 863 Projects of China (2008AA01A323 and 2008AA01A326), and Program for New Century Excellent Talents in University.

References

[1] A. S. Tanenbaum, *Computer Networks*, 4th ed. Upper Saddle River, USA: Prentice Hall, 2006.
[2] G. Hardin, "The Tragedy of the Commons," *Science*, vol. 162, no. 3859, pp. 1243 – 1248, Dec. 1968.
[3] R. Cocchi, D. Estrin, S. Shenker, and L. Zhang, "A Study of Priority Pricing in Multiple Service Class Networks," in *Proc. conf. Communications Architecture & Protocols (SIGCOMM)*, Zurich, Switzerland, pp. 123 – 130, 1991.
[4] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang, "Pricing in computer networks: motivation, formulation, and example," *IEEE/ACM Trans. Netw.*, vol. 1, no. 6, pp. 614 – 627, Dec. 1993.
[5] S. Androutsellis-Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies," *ACM Comput. Surv. (CSUS)*, vol. 36, no. 4, pp. 335 – 371, Dec. 2004.
[6] Akamai, http://www.akamai.com.
[7] M. Welzl, *Network Congestion Control: Managing Internet Traffic*. Chichester, England: John Wiley & Sons.
[8] S. H. Low, "A duality model of TCP and queue management algorithms," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 525–536, Aug. 2003.
[9] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556 – 567, Oct. 2000.
[10] N. Wang, K. Ho, G. Pavlou, and M. Howarth, "An overview of routing optimization for internet traffic engineering," *IEEE Commun. Surv. Tutor.*, vol. 10, no. 1, pp. 36 – 56, 2008.
[11] V. Valancius, N. Laoutaris, L. Massoulie, C. Diot, and P. Rodriguez, "Greening the Internet with nano data centers," in *Proc. 5th Int. conf. Emerging Networking Experiments and Technologies (ACM CoNEXT)*, Rome, Italy, 2009, pp. 23 – 48.
[12] L. Qiu, Y. Yang, Y. Zhang, and S. Shenker, "On selfish routing in internet-like environments," in *Proc. ACM SIGCOMM*, Karlsruhe, Germany, 2003, pp. 151 – 162.

[13] R. Braden, D. Clark, and S. Shenker, “RFC 1633: Integrated services in the Internet architecture: an overview,” IETF, Jun. 1994.

[14] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “RFC 2475: An Architecture for Differential Services,” IETF, Dec. 1998.

[15] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala. “RSVP: a new resource reservation protocol.” *IEEE Network*, vol. 7, no. 5, pp. 8 – 18, Sept. 1993.

[16] D. D. Clark, “A model for cost allocation and pricing in the Internet,” Technical report, MIT, Aug. 1995.

[17] D. D. Clark, “Internet cost allocation and pricing,” in *Internet Economics*, L. W. McKnight and J. P. Bailey, Eds. Cambridge, USA: MIT Press, 1997, pp. 216 – 252.

[18] J. MacKie-Mason and H. Varian, “Pricing the Internet,” in *Public Access to the Internet*, B. Kahin and J. Keller, Eds. Cambridge, USA: MIT Press, 1995, pp. 269 – 314.

[19] J. MacKie-Mason and H. Varian, “Pricing congestible network resources,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1141 – 1149, Sept. 1995.

[20] R. J. Edell, N. McKeown, and P. P. Variya, “Billing users and pricing for TCP,” *IEEE J. Sel. Areas Commun.*, Vol. 13, No. 7, pp. 1162 – 1175, Sept.1995.

[21] M. Currence, A. Kurzon, D. Smud, and L. Trias, “A Causal analysis of usage-based billing on IP networks,” University of Colorado, 2000. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.2035

[22] M. K. Honig and K. Steiglitz, “Usage-based pricing of packet data generated by a heterogeneous user population,” in *Proc. IEEE INFOCOM 1995*, Boston, USA, pp. 867 – 874.

[23] R. J. Gibbens and F. P. Kelly, “Resource pricing and the evolution of congestion control,” *Automatica*, vol. 35, no. 6, pp. 1969 – 1985, Dec. 1999.

[24] F. P. Kelly, “Charging and rate control for elastic traffic,” *Eur. Trans Telecomm.*, vol. 8, no. 1, pp. 33 – 37, Jan. 1997.

[25] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, “Rate control in communication networks: shadow prices, proportional fairness and stability,” *J. Operational Research Society*, vol. 49, no. 3, pp. 237 – 252, Mar. 1998.

[26] J. Crowcroft and P. Oechslin, “Differentiated end-to-end Internet services using a weighted proportionally fair sharing TCP,” *ACM Comput. Commun. Rev.*, vol. 28, no. 3, pp. 53 – 67, Jul. 1998.

[27] S. Kunniyur and R. Srikant, “End-to-end congestion control: utility functions, random losses and ECN marks,” in *Proc. IEEE INFOCOM 2000*, Tel Aviv, Israel, pp. 1323 – 1332.

[28] P. Milgrom, *Putting Auction Theory to Work.* Cambridge, UK: Cambridge University Press, 2004.

[29] J. MacKie-Mason, “A smart market for resource reservation in a multiple quality of service information network,” Technical report, University of Michigan, Sept. 1997.

[30] N. Keon and G. Anandalingam, “A new pricing model for competitive telecommunications services using congestion discounts,” *INFORMS J. Comput.*, vol. 17, no. 4, pp. 248 – 262, Apr. 2005.

[31] M. Yuksel and S. Kalyanarama, “Pricing granularity for congestion-sensitive pricing,” in *Proc. Computers and Communications (ISCC 2003)*, Antalya, Turkey, vol. 1, pp. 169 – 174.

[32] R. J. Edell and P. Varaiya, “Providing Internet Access: What we Learn from the INDEX Trial,” INDEX Project Report 99-010W, Apr. 1999.

[33] S. Shenker, “Some fundamental design decisions for the future Internet,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176 – 1188, Sept. 1995.

[34] J. W. Roberts, “Quality of service guarantees and charging in multi-service networks,” *IEICE Trans. Commun.*, vol. 81-B, no. 5, pp. 824 – 831, May 1998.

[35] M. Odlyzko, “Paris metro pricing for the internet,” in *Proc. 1st ACM conf. Electronic Commerce*, Denver, USA, 1999, pp. 140 – 147.

[36] K. Nichols, V. Jacobson, and L. Zhang, “RFC 2638: A two-bit differentiated services architecture for the Internet,” Internet request for comments, IETF, Jul. 1998.

[37] S. Kalyanaraman, T. Ravichandran, and R. Norsworthy, “Dynamic capacity contracting: a framework for pricing the differentiated services Internet,” in *Proc. 10th Annu. Workshop Information Technologies and Systems (WITS)*, Brisbane, Australia, 2000.

[38] M. Yuksel and S. Kalyanaraman, “Distributed Dynamic Capacity Contracting: An overlay congestion pricing framework,” in *Proc. Computers and Communications (ISCC 2003)*, Antalya, Turkey, 2003, vol. 26, pp. 1484 – 1503.

[39] M. Karsten, J. Schmitt, L. Woff, and R. Steinmetz, “An embedded charging approach for RSVP,” presented at *Int. Workshop Quality of Service*, Napa, USA, May 1998.

[40] D. D. Clark, “Combining sender and receiver payments in the Internet,” in *Interconnection and the Internet: Selected Papers from the 1996 Telecommunications Research Policy Conference*, G. L. Rosston and D. Waterman, Eds. Mahwah, USA: Lawrence Erlbaum Associates, 1997, pp. 95 – 112.

[41] G. Fankhauser, B. Stiller, C. Vögtli, and B. Plattner, “Reservation-based charging in an integrated services network,” in *Proc. 4th INFORMS Telecommunications Conf.*, Boca Raton, USA, 1998.

[42] G. Fankhauser and B. Plattner, “Diffserv bandwidth brokers as mini-markets,” in *Int. Workshop Internet Service Quality Economics*, Cambridge, USA, 1999.

[43] N. Semret, R. R.-F. Liao, A. T. Campbell, and A. A. Lazar, “Pricing, provisioning and peering: Dynamic markets for differentiated internet services and implications for network interconnections,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 12, pp. 2499 – 2513, Dec. 2000.

[44] X. Wang and H. Schulzrinne, “An integrated resource negotiation, pricing, and QoS adaptation framework for multi-media applications,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 12, pp. 2514 – 2529, Dec. 2000.

[45] X. Wang and H. Schulzrinne, “Pricing network resources for adaptive applications in a differentiated services network,” in *Proc. IEEE INFOCOM 2001*, Anchorage, USA, vol. 2, pp. 943 – 952.

[46] A. Gupta, D. Stahl, and A. Whinston, “Priority pricing of integrated services,” in *Internet Economics*, L. W. McKnight and J. P. Bailey, Eds. Cambridge, USA: MIT Press, 1997, pp. 323 – 352.

[47] A. Gupta, D. O. Stahl, and A. B. Whinston, "An economic approach to network computing with priority classes," *J. Organizational Computing Electronic Commerce*, vol. 6, no. 1, pp. 71 – 95, 1996.

[48] M. H. Dahshan and P. K. Verma, "Resource based pricing framework for integrated services networks," *J. Networks*, vol. 2, no. 3, pp. 36 – 45, Jun. 2007.

[49] C. Dovrolis and P. Ramanathan, "A case for relative differentiated services and the proportional differentiation model," *IEEE Network*, vol. 13, no. 5, pp. 26 – 34, Oct. 1999.

[50] C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional differentiated services: Delay differentiation and packet scheduling," *IEEE/ACM Trans. Netw.*, vol. 10, no. 1, pp. 12 – 26, Feb. 2002.

[51] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, USA: Athena Scientific, 1999.

[52] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, no. 2, pp. 289 – 295, Sept. 1951.

[53] X. R. Cao, "Preference functions and bargaining solutions," in *Proc. 21st IEEE Conf. on Decision and Control*, Orlando, USA, 1982, pp. 164 – 171.

[54] M. Simaan and J. B. Cruz, Jr, "On the Stackelberg strategy in nonzero-sum games," *J. Optim. Theory Appl*, vol. 11, no. 5, pp. 533 – 555, Sept. 1973.

[55] T. Basar and R. Srikant, "A Stackelberg network game with a large number of followers," *J. Optim. Theory Appl.*, vol. 115, no. 3, pp. 479 – 490, Dec. 2002.

[56] T. Basar and R. Srikant, "Revenue-maximizing pricing and capacity expansion in a many-users regime," in *Proc. IEEE INFOCOM 2002*, New York, USA, vol. 1, pp. 294 – 301.

[57] H. X. Shen and T. Basar, "Differentiated Internet pricing using a hierarchical network game model," in *Proc. American Control Conf.*, Boston, USA , 2004, pp. 2322 – 2327.

[58] H. X. Shen and T. Basar, "Optimal nonlinear pricing for a monopolistic network service provider with complete and incomplete information," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp.1216 – 1223, Aug. 2007.

[59] E. Altman, D. Barman, R. E. Azouzi, D. Ros, and B. Tuffin, "Pricing differentiated services: A game-theoretic approach," *Comput, Netw.*, vol. 50, no. 7, pp. 982 – 1002, May 2006.

[60] D. Acemoglu and A. Ozdaglar, "Competition and Efficiency in Congested Markets," *Mathematics of Operations Research*, vol. 32, no. 1 pp. 1 – 31, Feb. 2007.

[61] J. F. Nash, Jr, "The bargaining problem," *Econometrica*, vol. 18, no. 2, pp. 155 – 162, Apr. 1950.

[62] A. Roth, *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge, UK: Cambridge University Press, 1988.

[63] X. R. Cao, H. X. Shen, R. Milito, and P. Wirth, "Internet pricing with a game theoretical approach: concepts and examples," *IEEE/ACM Trans. Netw.*, vol. 10, no. 2, pp. 208 – 216, Apr. 2002.

[64] H. Yaiche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 667 – 678, Oct. 2000.

[65] R. T. Ma, D. M. Chiu, J. C. S. Lui, V. Misra, and D. Rubenstein, "On cooperative settlement between content, transit and eyeball internet service providers," in *Proc. 5th Int. Conf. Emerging Networking Experiments and Technologies (ACM CoNEXT)*, Madrid, Spain, 2008.

[66] R. T. Ma, D. M. Chiu, J. C. S. Lui, V. Misra, and D. Rubenstein, "Interconnecting eyeballs to content: A Shapley value perspective on ISP peering and settlement," in *Proc. ACM Network Economics (NetEcon)*, Seattle, USA, 2008, pp. 61 – 66.

[67] J. Altmanna and K. Chu, "How to charge for network services: flat-rate or usage-based?" *Comput. Netw.*, vol. 36, no. 5 – 6, pp. 519 – 531, Aug. 2001.

[68] http://www.cernet.edu.cn/20010912/3001298.shtml

[69] C. Courcoubetis, G. D. Stamoulis, C. Manolakis, and F. P. Kelly, "An intelligent agent for optimizing QoS-for-money in priced ABR connections," in *Proc. Int. Conf. Telecommunications (ICT)*, Halkidiki, Greece, 1998.

[70] Q. Wang, D. M. Chiu, and J. C. S. Lui, "ISP Uplink Pricing in a Competitive Market," in *Proc. Int. Conf. Telecommunications (ICT)*, St. Petersburg , Russia, 2008, pp. 1 – 6.

[71] H. He and K. Xu, "Can P2P technology benefit ISPs? A cooperative profit-distribution answer," Technical Report, Department of Computer Science and Technology, Tsinghua University, Jul. 2011.

[72] N. Nisan and A. Ronen, "Algorithmic mechanism design," in *Proc. 31st Annu. ACM Symp. Theory of Computing*, Atlanta, USA, 1999, pp. 129 – 140.

[73] S. Shenker, D. Clark, D. Estrin, and S. Herzog, "Pricing in computer networks: reshaping the research agenda," *ACM Comput. Commun. Rev.*, vol. 26, no. 2, pp. 19 – 43, Apr. 1996.

[74] P. Dube, V. Borkar, and D. Manjunath, "Differential join prices for parallel queues: social optimality, dynamic pricing algorithms and application to Internet pricing," in *Proc. INFOCOM 2002*, New York, USA, vol. 1, pp. 276 – 283.

[75] A. J. O'Donnell and H. Sethu, "Congestion control, differentiated services, and efficient capacity management through a novel pricing strategy," *Comput. Comm.*, vol. 26, no. 13, pp. 1457 – 1469, Aug. 2003.

[76] T. Li, Y. Iraqi, and R. Boutaba, "Pricing and admission control for QoS-enabled Internet," *Comput. Netw.*, vol. 46, no. 1, pp. 87 – 110. Sept. 2004.

[77] M. N. Gregory, *Principles of Economics*, 5th ed. Translated by X. M. Liang and L. Liang. Beijing, China: Peking University Press, 2009.

[78] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. Netw.*, vol. 10, no. 2, pp. 272 – 286, Apr. 2002.

[79] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Non-convex optimization and distributed pricing based algorithms for optimal resource allocation in high speed networks," in *Proc. 17th IEEE Annu. Computer Communications Workshop*, Santa Fe, USA, 2002.

[80] J. W. Lee, R. R. Masumdar, and N. B. Shroff, "Non-convex optimization and rate control for multi-class services in the internet," *IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 827 – 840, Aug. 2005.

[81] S. Stidham, Jr, "Pricing and congestion management in a network with heterogeneous users," *IEEE Trans. Autom. Control*, vol. 49, no. 6, pp. 976 – 981, Jun. 2004.

[82] M. Chiang, S. Zhang, and P. Hande, "Distributed rate allocation for inelastic flows: Optimization frameworks optimality condition, and optimal algorithms," in *Proc. IEEE INFOCOM 2005*, Miami, USA, vol. 4 pp. 2679 – 2690.

[83] P. Hande, S. Rangan, and M. Chiang, "Distributed algorithms for optimal SIR assignment in cellular data networks," in *Proc. IEEE INFOCOM 2006*.

[84] P. Hande, S. Zhang, and M. Chiang, "Distributed rate allocation for inelastic flows," *IEEE/ACM Trans. Netw.*, vol. 15, no. 6, pp. 1240 – 1253, Dec. 2007.

[85] P. Hande, M. Chiang, R. Calderbank, and J. Zhang, "Pricing under constraints in access networks: revenue maximization and congestion management" in *Proc. IEEE INFOCOM 2010*, San Diego, USA, pp. 1 – 9.

[86] P. Hande, M. Chiang, A. R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," in *Proc. IEEE INFOCOM 2009*, Rio de Janeiro, Brazil, pp. 990 – 998.

[87] R. Beverly, S. Bauer, and A. Berger, "The internet's not a big truck: Toward quantifying network neutrality," in *Proc. Passive and Active Measurements (PAM) Workshop*, Louvain-la-neuve, Belgium, 2007, pp. 135 – 144.

[88] R. B. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, USA: Harvard University Press, 1991.

[89] S. Shakkottai, R. Srikant, A. Ozdaglar, and D. Acemoglu, "The price of simplicity," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1269 – 1276, 2008.

[90] S. Li and J. Huang, "Revenue maximization for communication networks with sage-based pricing," in *Proc. IEEE GLOBECOM*, Hawaii, USA, 2009.

[91] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407 – 435, Aug. 2004.

[92] R. Johari, "Efficiency loss in market mechanisms for resource allocation," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, USA, 2004.

[93] A. Ozdaglar and R. Srikant, "Incentives and pricing in communication networks," in *Algorithmic Game Theory*. Cambridge, USA: Cambridge University Press, 2007, ch. 22, pp. 571 – 591.

[94] S. C. Lee, J. W. Jiang, D. M. Chiu, and J. C. S. Lui, "Interaction of ISPs: Distributed Resource Allocation and Revenue Maximization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 2, pp. 204 – 218, Feb. 2008.

[95] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulie, "Incentivizing peer-assisted services: a fluid shapley value approach," in *Proc. SIGMETRICS*, New York, USA, 2010, pp. 215 – 226.