Modelling and Analysis of Internet Pricing and Revenue Distribution

CHEUNG, Yang

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Information Engineering

Abstract of thesis entitled:

Modelling and Analysis of Internet Pricing and Revenue Distribution

Submitted by CHEUNG, Yang

for the degree of Master of Philosophy

at The Chinese University of Hong Kong in September 2008

The economic revenue model in the Internet is to collect revenues from users and distribute them among network players. The current collecting and distributing mechanism in the Internet has introduced network inefficiency and unfairness, which leads to the lack of incentive for further Internet investment.

Our contribution of this work is to address two important aspects of the economic revenue model in the future Internet. First, we address the network inefficiency problems raised by the current Internet in the *revenue collection stage*. We analyzed how the concept of insurance can be adopted into Internet pricing model to users. The insurance we analyze is namely price stability insurance, which aims to alleviate price fluctuations by congestion pricing by bearing a portion of users' congestion charges.

We discussed the responses of price stability insurance from users with varies risk adversities. We proposed a coinsurance function such that users can enjoy minimum price fluctuations while maintaining incentive for them to fine tune their network behavior.

Second, we address the fairness issues in the revenue distribution stage. We determine the existence and realizability of bi-

lateral prices that can achieve fair revenue division among ISPs. In particular, we use Shapley value as the basis for deriving fair prices. Under a quite general topology and traffic model, we find that there exists prices that make the revenue division under bilateral settlement equal to that calculated under Shapley value. The corresponding "fair price" exhibits several nice and desirable characteristics. Moreover, it could be realized approximately.

在互聯網的經濟模式之下，網絡供應商向用戶收取收益，再分配與各個網絡上的參與者。現存網絡收益的收集及分配機制引致網絡運作欠缺效率及造成不公平的現象，進一步使投資者缺乏在互聯網上投資的意欲。

這份論文從兩個大角度去回應未來互聯網的經濟模式。首先，我們針對現今互聯網的收益收集過程所導致的欠缺網絡效率，分析如何能夠將保險的概念納入互聯網的定價模式之中。我們將介紹一種價格穩定保險，當用戶造成網絡擠塞而須要繳付額外收費時，透過提供優惠而緩和由網絡擠塞所造成的價格波動。

我們將討論不同風險承受程度的用戶對於上述保險的回應。從中我們建議以一個折扣率算式，使用戶能夠在享受最小的價格波動的同時，保持他們改善其網絡行為的意欲。

其次，我們就網絡供應商在收益分配中的不公平情況作出分析。我們證實透過相互價格達致公平地在網絡供應商之間分配收益的方法是存在及可行的。針對而言，我們利用斯克理價格作為衍生公平價格的基礎。在一個大致基本的拓撲及網絡交通模式中，我們證實一個與斯克理價格相同的相互價格在收益分配上是存在的，而這一個「公平」的相互價格亦展示出良好及適合互聯網的特性。再者，這個公平的相互價格是能夠在缺乏完整資訊下實現的。

Acknowledgement

Writing this thesis brings me much challenges. Without the inspiration, guidance and support of the following people, this study would not have been completed. I would like to use this opportunity to express my deepest gratitude to them.

- Professor Chiu Dah Ming who brought me to an eye-open journey of the world of research. His wisdom, his enthusiasm and insight on research problems have established a perfect role model for me. He is the reason why I want to pursue further in academics.

- Professor Huang Jianwei who motivated me during the uncertain times of the research process. His persistence to correctness, his patience, and his supprot have deeply impressed me. His research attitude is what I need to learn by heart.

- My family, who has always supported me and believed in me even at times when I do not believe in myself.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the bird-eye view of the Internet, network players, such as Internet Service Providers(ISPs), are sitting arbitrary in the Internet cloud. They are interconnected with each other to establish connectivity. Users are connected at the edge of the Internet cloud by edge ISPs, which provide connectivity for them in exchange for a revenue. The economic revenue model in the Internet is simple: In the *revenue collecting stage*, edge ISPs collect revenues from users asking for connectivity; In the *revenue distributing stage*, revenues are passed among network players in the Internet.

The current economic revenue model in the Internet is becoming more problematic with the advancement of network technologies and the change of users' behaviors. In the *revenue collecting stage*, users' behaviors change with the popularity of high speed broadband access. Their reckless use of network resources give much burden to the network capacity, nonetheless users do not have the correct incentive to make the network efficient; In the *revenue distributing stage*, the division of revenue heavily depends on external factors such as ISP's negotiating power and their market shares, without regarding the fundamental contribution of ISPs. We can foresee that the Internet is incubating a hostile environment to both users and network players, where the Internet would eventually become too wasteful and inequitable to encourage further expansion or investment.

In this paper, we would like to answer this question: how can we design a proper economic revenue model for the future Internet? Although this question is board and complicated to be analyzed thoroughly, we would like to address two key parts of the question.

First, we focus on the *revenue collecting stage* of the economic revenue model. There are vast amount of literatures on pricing mechanisms that can help improving network efficiency in the Internet. However, those mechanisms are often unacceptable to users since they are either too complicated or the pricing structure is uncomfortable to users. In our work, we help novel pricing mechanisms from literatures to become feasible to implement by introducing the concept of insurance. The introduction of insurance is beneficial in twofold: Not only it can reduce price fluctuations and data rate turbulences experienced by users, but also help ISPs to obtain a guaranteed return of investment under pricing mechanisms such as congestion pricing.

Second, we focus on the *revenue distributing stage* of the economic revenue model. The revenue distribution among ISPs in the Internet currently are non-cooperative and unfair. In our work, we discuss how network-wide cooperative settlement can be reached by using normal bilateral settlement method currently adopted in the Internet. As a fair criteria, we use Shapley value as the standard of fairness to determine the fair price obtained by each ISP. We also showed how this network-wide cooperative price can be obtained in an asymmetric environment with limited information.

This paper is organized as follows: In Chapter 2, we present the literatures related to economic revenue model. In Chapter 3, we focus on the *revenue collecting stage* of the economic revenue model, where we discuss how insurance can be applied to Internet pricing to make novel pricing mechanisms feasible to both

ISPs and users. In Chapter 4, we focus on the *revenue distributing stage* of the economic revenue model, where we discuss how network-wide cooperative settlement can be achieved by current bilateral settlements. We conclude this work in Chapter 5.

□ **End of chapter.**

Chapter 2

Related Works

2.1 Pricing Mechanisms

2.1.1 Current Situation

Flat rate pricing [1] is the dominating pricing mechanism between ISPs and users. Flat rate pricing is to charge users a fixed charge for unlimited access to the Internet in a period of time, for example a month. Depends on how much a user pays, different upper bandwidth bounds are applied. There are no lower bandwidth bound for common users. This charging mechanism is simple to be implemented by ISPs as no monitoring system is required. in users' perspective, this mechanism is intuitive, and this completely predictable charging makes generally risk-averse users comfortable. However, flat rate pricing does not advocate efficient use of the Internet. Since flat rate pricing charges users the same amount whatever they use the network, users tend to use the network recklessly and this causes network inefficiency.

The phenomenon of network inefficiency is getting more severe with the increase of last mile bandwidth. With broadband access, users do not limit their network applications to emails and web browsing. They are able to run inelastic applications such as video streaming, or high traffic applications such as peer-to-peer data transfer programs. A white paper from MIT

Communications Future Group [36] mentioned that the behavior change of users due to the introduction of broadband under flat rate pricing is one of the reasons of network inefficiency, which prevents ISPs to increase their network capacities. Together with the traffic per user increases by the change of users' behaviors, the operating cost of ISPs increases. The revenue obtained by users, however, is not able to catch up with the increase of traffic under flat rate pricing. Normally 20% of the users contribute to 80% of the traffic, therefore, it is unreasonable to charge flat rate to every user in proportion to network's total traffic as that will make 80% of the users unhappy.

Odlyzko [34] discussed the issue of whether flat rate pricing should be replaced by a more efficient pricing mechanism in a practical point of view. He argued that simplicity should win over efficiency. Since users do not mind paying more for a simpler mechanism, implementing a simple mechanism is a win-win solution to both ISPs and users. While we agree that a simple pricing is vital for users' adaptation, we still need to overcome the huge amount of fluctuating data rates in the future Internet, which is not possible to be solved purely by flat rate pricing and over-provisioning.

Currently, there are several solutions to tackle the increasing data rate for users and the decreasing of ISPs' per-unit bandwidth profit. One method is to throttle users' traffic by setting an upper bandwidth bound. This method is pragmatic, however, is not reasonable when the network is not congested. In our point of view, the network should be classified as a common good. When the network resource is non-excludable, that is ones usage of the network does not prevent others from using the network, one should pay nothing for the network resource. Besides, throttling users' traffic also violates network neutrality [44] as applications that require bandwidth higher than the upper bandwidth bound is strictly prohibited to the user because of the network restriction. Another method to increase ISPs' profit is to provide service differentiation or content differentiation. By paying a higher price, users can enjoy a better quality of service, or can use some addition applications provided by the ISPs.

An alternate solution is to use other pricing mechanisms proposed by varies researchers instead of flat rate pricing, which are presented in the following context.

2.1.2 Proposed Pricing Mechanisms

The pricing mechanism proposed currently can generally divided into four categories. The first type is fixed price charging, which flat rates are charged regardless of the users' network usage behaviors. The second type is traffic value based charging, which the charge of the network usage is reflected by how much a user value ones traffic. The third type is usage based charging, which the charge is proportional the actual traffic injected into the network. The forth type is congestion charging, which there are only charges applied to users when they initiate traffic when the network is congested. Detailed survey paper introducing varies pricing mechanisms can be found in [37, 12]

In fixed price charging ISPs set a fixed price by certain criteria. Users subscribed to those ISPs pay the deterministic price for whatever bandwidth they use. Examples of such pricing scheme are flat rate pricing, Paris metro pricing [33] and edge pricing [40]. The concept of Paris metro pricing is to divide network capacity into several equal bandwidth sub-channels, and to charge each sub-channel at a different flat rates. Using this setting, users with important applications switch to a more expensive sub-channel. This helps ISPs to reveal users' valuation of their traffic and make according charging. Edge pricing is to divide prices with respect to time rather than bandwidth in

Paris metro pricing. It is also referred as time-of-day pricing, as users with varies time-based usage behaviors are charged differently. For instance, the charge for users who need to use the network during working hours are higher than users who only need network at off-peak hours. Therefore, we can expect that users who care about the additional charges at peak times shift their traffic to times when the rate is small. Such shifting of traffic helps reducing bandwidth fluctuation in the network across time. Fixed price charging is ideal in terms of compatibility to the network, low accounting cost and user-friendly. It, however, is well known to be ineffective [11] because these pricing mechanisms does not limit users' usage, thus there are no incentive for users to use the network responsibly.

Traffic-value based pricing is to ask users to give a value to their traffic, so that the network can obtain necessary information to perform optimal network resource allocation and charging. Examples of such pricing are smart-market pricing [30] and priority pricing [16]. Smart-market pricing is inspired by auctions. In smart-market pricing, each traffic issued by a user carries a bid with it, and the data rate allocation is an auction based mechanism. A number of traffic with the highest bids win and bandwidth is allocated to them. Act as an incentive to prevent reckless bidding, a charge proportional to the bid is needed to pay for the traffic. This charging scheme provides good efficiency as it is able to allocate resource to the most needy applications; however, building an auction system is complicated and often centralized. Priority pricing is on the other hand inspired by the priority mailing service. Similarly, priority pricing asks users to select a priority level of the traffic. The network then selects the ones with higher priority to be first delivered. Of course the higher the priority the traffic is, the more expensive the traffic charge is. One major drawback of traffic-value based traffic is the frequent decision making. Users are required to make judgements on the value (or class) of each of their traffic. This decision frequency not only infeasible in practice, but also users are reluctant to making such frequent decisions.

Usage based charging treats network bandwidth as commodities which users pay according to their usage. Example of such are q-th percentile pricing, expected capacity pricing [9] and effective bandwidth pricing [22]. In q-th percentile pricing, bandwidth usage of a user in a period is divided the into several discrete time slots. The bandwidth usage charged to an user is determined by at the q-th percentile time slot. The meaning of charging this way is to ask users to take responsibility to their peak data rates for ISP's bandwidth reservation use, rather than their overall bandwidth usage. Expected capacity pricing is to charge users by an estimation of the expected capacity rather than the actual bandwidth usage. Effective bandwidth pricing uses a mathematical approach to make sure the user declares the accurate mean and the deviation of their traffic pattern. Usage based charging gives incentive to users to better use the network since charges are incurred in each traffic. It does not, however, provide enough incentive for users to adjust their usage at critical scenarios such as during congestion.

Congestion pricing is to treat network as a common good when the network is not congested, and charge according to the congestion situation when the network turns into a private good. Example of this pricing mechanism are responsive pricing [29] and smart-market pricing [30]. In responsive pricing, the per-unit charge is proportional to the congestion index. At each time slot, network announces the congestion index to all users, users then respond with their optimized data rate according to the congestion index. This charging mechanism correctly identifies the properties of network capacity and gives correct incentive for users to adjust their usage during congestion. This mechanism, however, does not guarantee return on investment for network providers. Also the charges for users are expected to experience fluctuations, which users are not feeling comfortable with.

2.1.3 Congestion Pricing

As part of this work is built mainly on congestion pricing, it worths more detail review on this topic. Early work related to congestion pricing is by Kelly [23]. He have successfully built a relationship between payment and data rate allocation by using the concept of shadow price. The shadow price is the dual variable in the lagrange multiplier pattern, which its literal meaning is the additional gain by increasing one unit of the constraint variable. In Kelly work, he introduced the concept of using shadow price as usage-based charges. By this formulation, shadow price provides appropriate signal for users to chose their respective data rates, such that users' optimization problem, network's optimization problem and social's optimization problem can be simultaneously satisfied.

The implementation of congestion pricing is discussed in Gibbens and Kelly [15]. It showed that by marking every packet when a resource is overloaded and charging each marked packet a price, the expected value of the total charge is precisely equals to the shadow price charging method stated previously. However, this result is only valid when the data rates across time are in Poisson distribution. Moreover, the work does not consider the situation where there are packet drops, either by noise or by congestion. In order for congestion pricing to be implemented in the network, we need a better settlement method.

For this issue, Briscoe *et al* proposed the concept of re-feedback [7] and re-ECN [5], which aims to feedback information such as congestion or delay information to the sender. One way of achieving this is to modify the time to live (TTL) counter in the TCP header. By adjusting the TTL counter at the sender to a smaller value than default (255), the received TTL information can be used as an indicator of the network situation. This work helps the right person in the transmission line(the sender) to know the network situation, and to obtain congestion information for the payment. One doubt is that common routers might treat TTL as hop count because time is rather difficult to calculate. If that is true, using TTL might not be a valid tool to indicate congestion/delay time. The drawback is that the mechanism is not incentive compatible. Malicious ISPs along the path can assert additional TTLs, so that the signal sent back to the sender appears more congested and a higher congestion charge is incurred to the user. Another drawback of Briscoe's proposal is that the last reserved bit in TCP header will be used in his re-ECN proposal.

To tackle the incentive compatible issue, Laskowski and Chuang [25] mentioned how to prevent cheating and coalition of ISPs by implementing a rest of path (ROP) monitor at each ISP. The function of a ROP monitor is to determine the quality for the rest of the path to the destination. The paper claims that network's innovation (such as incentive to increase capacity) and accountability is correlated. By making ISPs accountable for their behaviors when ROP monitors are implemented, innovations can be encouraged. As a digressed remark, implementing ROP monitors still does not provide enough incentive for ISPs to undergo end-to-end innovations such as increasing capacities, as their sole work of innovation does not provide any difference when the rest of path does not innovate.

2.1.4 Bandwidth Allocation Mechanism

In users' perspective, their main goal is to maximize their utility surplus, which is the difference between the utility obtained from receiving certain data rate and the cost of obtaining the data rate. It is, therefore vital for us to understand how network allocates data rates in order to better acknowledge the users' decisions under different data rate allocation schemes. A detailed review of fairness issues in bandwidth allocation could be found in the writing of Le Boudec [26].

The currently implemented data rate allocation method is max-min fair [4]. It is as simple as dividing the data rates evenly to all participated flows. It is easy to implement by using additive increase multiplicative decrease algorithm [8], however, it is known that max-min fair mechanism is not fair to large flows, as more bandwidth could be obtained through small flows in proportion than in large flows. As an alterative, Kelly *et al* proposed proportional fairness [23, 24]. In proportional fairness, the bandwidth is allocated in proportion to the requested data rate. Weighted proportional fairness is also introduce by Kelly [23] that the bandwidth is determined by both the requested data rate and a weight factor for each flow.

In weighted proportional fairness approach, users intend to transmit traffic sent out a value indicating their expected payment as the weight factor to the network. The network then calculates the data rates allocated to each user. Hajek and Yang [17] showed that such mechanism is not strategy-proof. When users are strategic rather than merely a price-taker, the network efficiency of such mechanism would be low. Therefore, Yang and Hajek [45] proposed VCG-Kelly mechanism, which uses VCG mechanism as a wrapper to the original Kelly's mechanism to make the allocation strategy-proof. However, since VCG mechanism is centralized in nature, which needs complete information of other users in order to find the second largest price, there are still works to be done to decentralize the mechanism, in addition to reduce the complexity of the mechanism.

The term "fair" in the previous context is about dividing bandwidth across different flows in an indifferent way. However, since each application is able to establish as many flows as it wish, the fairness criteria cannot stand when we view fairness in application level. In light of this, Briscoe[6] proposed cost fairness to tackle this issue. He mentioned that instead of allocating data rates among flows, network should allocate congestion by congestion volume among users. This concept is similar to congestion pricing, where each user can send out as many traffic as they wish, but if one traffic causes any congestion, the user is required to pay the congestion volume incurred by oneself. By using commercial means, flow rate can allocated under the invisible hand from Adam Smith.

2.2 Revenue Distribution Mechanisms

2.2.1 Current Situation

Settlements between ISPs are generally divided into two categories, namely provider-customer settlements and peer-to-peer settlements. For provider-customer settlements, q-th percentile pricing is the common practice. The charging mechanism is composed of two elements. The first element is the promised bandwidth charge, which the charge is proportional to minimum bandwidth promised. This charge indicates the minimum bandwidth a customer can get at any time guaranteed by the ISP, on the other hand, this charge also indicates the minimum bandwidth the customer promises to pay, whatever he actually use it or not in the coming period. The second element is the tile-based charge. Within a period (such as a month), the provider keeps record of the bandwidth used by each customer at each time interval (such as every five minutes). At the end of the period, each customer's bandwidth record is sorted, and the qth highest tile record is picked up as the amount of bandwidth that the provider has reserved for the customer for this period.

The actual tile-based charge is a function of the q-th highest tile and the promised bandwidth charge, which the function increases with increasing q-th highest tile and decreases with increasing promised bandwidth charge. This function advocates the customers to better estimate their promised bandwidth, in order to give providers a more stable income. An analysis of q-th percentile pricing is studied by Levy *et al*[27], which it discussed responses of q-th percentile pricing in several scenarios, and how the decision of multi-homing is drawn. For peer-to-peer settlements, senders-keep-all mechanism is usually adopted [20]. There are no monetary settlement between two peered ISPs for traffic to either direction.

As we can see, finding the optimal provider-customer relationship and peering relationship is vital to ISPs' profit in the current setting. Weiss and Shin [42] demonstrated the peering decision depends on market share of the ISP. By using a simple model, it showed that the two ISPs would establish a peering relationship if their market share is comparable such that their traffic at both sides are high enough to make the peering link profitable by both entities. However, in reality, the peering practice is much more complicated than profit. As stated by Norton [32], the peering decision depends on series of political, market power, company strategy, and many other concerns.

2.2.2 Novel Revenue Distribution Mechanisms

There are several literatures proposing new revenue distribution mechanisms. One way to distribute revenues among ISPs fairly is to use economic theories to do the distribution. Ma *et al* [28] analyzed how Shapley value can be implemented as a distribution mechanism in the Internet. By using Shapley value, the revenue obtained by each ISP is determined by its contribution to the connectivity of all traffic in the network. When revenue is distributed by Shapley value, several axioms such as additivity, dummy, symmetric and efficiency are automatically fulfilled. The work showed that a Nash equilibrium can be achieved when such profit-sharing mechanism is adopted. Also, when Shapley mechanism is implemented, optimal routing is the best response of all network players. Although the outcome of this mechanism is quite ideal, the work does not consider practical issues, such as implementation concerns of the mechanism, and the incentive of doing the change. The centralized Shapley value calculation is also another problem that worths further research.

He and Walrand [18] studies that if the revenue collected by ISPs are determined by their marginal demand on network capacity plus the traffic's lagrange multiplier (congestion shadow price), ISPs with smaller capacities obtain greater revenue in a non-cooperative setting. The analysis asserted that the revenue of each ISP is fair when it is proportion to ISP's bandwidth cost. By using weight proportional fairness criteria with respect to an ISP's cost, the ISP could obtain higher profit by collaborating. The drawback of this mechanism is that the proportional price requires global cost information. If there are only local information, the mechanism experiences efficiency loss. Also the paper based on a strong assumption that the higher the cost means the higher the investment, therefore the higher the negotiation power to gain more revenue. In reality, cost cannot directly reflect negotiation power, as it may be incurred by environment issues, or by poor management for example.

Accountability is always an issue in revenue distribution, since some network players may act malicious and lie about their cost in order to gain more. Feigenbaum *et al* [13] studied the accountability issue and proposed a distributed BGP-based routing mechanism that is strategy-proof. The concept is to use VCG mechanism to determine the charge of the intermediate ISPs along the traffic path. By paying intermediate ISPs the price of the alternate route minus their bandwidth cost, we can make sure that the best response to the intermediate ISPs are to indicate their true cost. However, this mechanism is not fair as the price charged by each ISP is greatly dependent on the cost of the alternate route, rather than the cost or the contribution of oneself. Another follow up question is the phenomenon of over-charging in VCG mechanism. It is possible that the sum of payment along the traffic path is much greater than the real cost of the path.

□ **End of chapter.**

Chapter 3

Problems in Revenue Collecting Stage

Summary

Currently, flat-rate pricing at access network in the Internet is commonly known for its network inefficiency. As a remedy, literatures have proposed several pricing mechanisms, such as congestion pricing. However, the implementation of those mechanisms is not feasible as they appose too much risk to both ISPs and network users. In this analysis, we adopted the concept of insurance to Internet pricing to mitigate the risk associated with those pricing mechanism. We also indicated how a price stability insurance can be set up in Internet pricing. Using those insurance products, we show preliminary results of how ISPs design the insurance and how users would respond to them accordingly.

3.1 Introduction

In this paper, we introduce the concept of insurance into Internet pricing. Similar to the concept of insurance in financial systems - where buyers are able to alleviate their monetary loss in the case of accidents - insurance in Internet pricing aims to help buyers from experiencing price fluctuations due to usage based Internet pricing mechanisms. We study the price stability insurance, which aims to mitigate price fluctuations resulted from the current novel Internet pricing mechanisms, such as congestion pricing [29]. By applying insurance to Internet pricing, we are able to make these Interne pricing mechanisms acceptable and feasible to be implemented in the perspective of network users.

The current traffic volume and pattern in the Internet has transformed as high speed access network becomes common. Nowadays, the Internet carries more inelastic and bulky applications than ever. We foresee that the Internet will become the carriage for every communication tool in the near future. The Internet pricing and data rate allocation mechanism, however, is lagging behind the rapid change of the Internet.

The current flat rate pricing in the Internet access network is more suited for the old days when Internet applications were mostly elastic and required little bandwidth. It was also appropriate as the Internet was trying to attract more users and Internet measurements were difficult to implement. Nowadays, however, the congestion-insensitive flat rate pricing is to blame for making the Internet suffer from network inefficiency. Flat rate charge does not provide incentive for network users to value their traffic, thus the network tends to be overused. This results in network congestions and decreased network efficiency. Moreover, the decrease of utility per capacity due to network inefficiency discourages ISPs to invest on additional capacity, which leads to a vicious circle of network congestion.

Although flat-rate pricing is are not welcomed by ISPs, it is preferable to users. Part of the reason is because of the risk adversity of common users. In general, users prefer stability and avoid variations. Flat-rate pricing promises a fixed rate no matter how much users consume. That helps users to minimize their risk on price fluctuation. Another part of the reason is the adversity of decision making. In flat-rate pricing, users only needs to ask one simple question once in a period - whether to subscribe or not. However, when it comes to usage based pricing, users need to make frequent decisions on how much traffic they send, or when they submit the traffic in order to minimize the amount they have to pay. The necessity for users to use the network tactfully make usage based pricing unattractive to users.

How to provide incentive for better usage of the Internet? Several novel Internet pricing mechanisms are among those being widely discussed. One of the examples is congestion pricing [29]. Congestion pricing implies that the user's payment is dependent upon how congested the network is: users pay nothing when the network is not congested; however, when the network is congested, users pay extra for the volume of congestion they add. The final payment by the user in the period is the sum of all congestion volumes incurred by oneself. Congestion pricing mechanism is feasible for implementation technically. The congestion situation can be observed by ISPs using packet drop rate, or by Explicit Congestion Notification (ECN) marks. ECN marks are asserted only when there are explicit congestion signals. ISPs can collect ECN marks by inspecting the TCP header of each packet to receive congestion signals. ISPs can charge users by charging each ECN mark returned from their traffic for some unit price. Nevertheless, the application of congestion pricing is not ideal in economics terms and users' habit.

There are three main flaws in congestion pricing. Firstly, congestion pricing provides temptation for ISPs to artificially increase network traffic. Since the payment from users increases with congestion, there is incentive for ISPs to keep their links congested at all times to incur more charges. Secondly, congestion pricing does not guarantee return on investment of ISP. As the payment from users fluctuates according to congestion, the ISP might not obtain sufficient revenue when the network is not congested. Thirdly is the users' adversities on price fluctuations and decision making as mentioned above.

3.1.1 Desirable Characteristics of Internet Pricing Mechanism

Before we mention the existing solution to tackle the flaws in congestion pricing, we would like to classify the important criteria that an ideal pricing mechanism should preserve in the future Internet. The criteria can be divided into three categories:

A) **Efficiency Issues**: the current flat rate pricing is known to be inefficient [12]. The ideal pricing mechanism should provide incentive for users to better utilize the Internet, including:

1. Increases network efficiency: similar to electricity, unused network capacity is wasted. The pricing mechanism should encourage users to use the network when it is not congested for better usage of network resources.

2. Increases economic efficiency: the pricing mechanism should increase social utility per capacity, so that resources are better allocated.

B) **User/ISP Adaptation Issues**: the human factor is one of the most important criteria in deciding whether the pricing mechanism is practical. The ideal pricing mechanism should take into account the behavior of the users, such as:

1. Incurs minimal fluctuation in users' payment - users should be able to estimate the amount they have to pay for that period in advance.

2. Avoids frequent decision making by users - users should only need to make decisions once in a while. They should not need to make microscopic fine-tuning of their network usage behaviors.

3. Guarantees Return on Investment - the total payment made by all users should be enough to cover the maintenance cost/building cost of the link regardless of the actual traffic, so that ISPs have enough motivation to establish links to any user;

C) **Social Issues**: being a common good, the Internet is vulnerable to tragedy of commons. Therefore, an ideal mechanism should help maintaining this common good status of the Internet, including:

1. Provides incentive for users to avoid congestion - the pricing mechanism should encourage users to reduce or minimize their traffic during congested time.

2. Avoids incentive for ISPs to artificially increase network traffic - ISPs should make no extra profit by artificially increase their network traffic. Instead, ISPs should increase their network capacities when congestion occurs.

When we consider congestion pricing for example, we can see that it satisfies the efficiency issues, but it fails in user/ISP practice issues and social issues. There is literature to suggest the social issues in congestion pricing are solved, as shown below.

3.1.2 Existing Solution

There have been attempts in literature to tackle the negative incentive of the ISPs to artificially increase traffic when the network is not congested. Anderson *et al.* [2] introduced a contract and balancing mechanism to solve the problems. The idea is that congestion cost are bared by the users, rather than between ISPs and users. Since the balancing process is performed amongst users rather than to ISPs, the amount of payments received by the ISP is fixed regardless of the network situation. Therefore, ISPs have no incentive to artificially increase network traffic.

In the mechanism, each user is contracted for a portion of the capacity from the ISP for a period of time. During the period, when the user generates more proportion of traffic than his contracted capacity, he is require to pay the additional traffic based on the congestion situation of the network. In contrast, when the user generates less than his contracted capacity, he is being paid by other users who have used the network excessively.

Nonetheless, this mechanism does not address the user habit issues. This mechanism implies additional risk of price fluctuation to users when congestion is not severe. Consider an extreme case when there is no traffic in the network, the user who tries to use the network would bear the entire capacity cost cY when one sends a traffic as one is reasonable for the entire proportion of traffic in the network. The fluctuation happens also when the users are not price-takers. If the users are being strategic, they would deliberately change their traffic pattern unpredictably and the resultant payment of users would vary further depending on other users' strategies.

In this study, we focus on the user habit issues based on the contract and balancing model mentioned above. We show that by applying the concept of insurance into Internet pricing, the fluctuation of payment of user can be mitigated. That helps solving the problems of congestion pricing and to make congestion pricing feasible for implementation in terms of economics and user habit issues.

3.1.3 Applying Insurance into Internet Pricing

Applying insurance into Internet pricing is one of the ways to help users get use to congestion pricing. The insurance scheme we proposed is named *price stability insurance*, that tackles the issue of price fluctuation in usage-based pricing mechanisms such as congestion pricing. The nature of price stability insurance follows the concept of coinsurance. In coinsurance, the insurer charges the buyer a certain amount of insurance ahead of time. When there are situations that the insurant needs to execute the insurance, the insurer covers a certain percentage of the cost, while the insurant bares the remaining portion. The percentage of coverage by the insurant is proportional to the amount of insurance purchased in advance.

Coinsurance is common in the field of health insurance [10]. In order to prevent the insurant from abusing the insurance, the insurance holder is obligated to pay for co-payment or coinsurance. Co-payment is to ask users to pay for a fixed price every time they use the health insurance. In coinsurance, the insurant pays a percentage of the total medical cost, while the rest is covered by the insurance. Analogue to health insurance, price stability insurance asks the insurant to pay a portion of the congestion charge the user incurred by transmitting traffic, while the rest is covered by the insurance.

Price fluctuation is not welcomed by network users as generally risk-averse users tend to feel more comfortable when they can estimate how much they need to pay in advance, even if they need to pay extra in exchange. The price stability insurance also avoids users from making frequent decisions - when price stability insurance is applied, users only need to decide the amount of insurance they have to buy once in a period, while for the rest of the time, users can be less cautious about the network condition as they are insured.

There are three main advantages of using insurance as a remedy to cushion price fluctuations for users. Firstly, the insurance mechanism is independent of the actual pricing mechanism used in Internet pricing. This does not only mean there requires no modification on the current Internet architectures and settings, the mechanism is also virtually transparent to ISPs. Secondly, the price stability insurance is based on a well established concept of insurance. The mechanism of the insurance is simple, the implementation is easy and well studied. Thirdly, the concept of insurance is not a stranger to most users. Users should have no problem in understanding the process and it is not difficult for users to find the mechanism reasonable.

In the mechanism of price stability insurance, a third-party insurance entity (the insurer) comes in and becomes the risk bearer. In the beginning of a period, the insurer provides a choice of insurance plans to individual network users. Users can choose to participate or not, or the amount of insurance they would like to purchase. During the period, users who purchase insurance can be coinsured when they incurs any network charges due to congestion until the end of a period, then the insurance subscription process restarts. At the end of the period, user pay a portion of the total congestion charge to the ISP, while the insurer bears the rest of the payment according to the insurance contract signed in the beginning of the period.

Establishing price stability insurance can help users solve the user habit issues in the following three ways:

Payment predictable to both network and users

When insurance is available, users are able to purchase insurance so that they incur less congestion charge when the network is congested. We can see that the more risk averse the user is, the more insurance the user will purchase. The effect of buying insurance is that the total payment can become more predictable. With higher the coinsurance coverage, the congestion charge demanded from ISP would become less significant. In an extreme case, a user would purchase enough insurance such that he/she was entitled to 100% coverage, the payment would be fixed for the entire period, similar to the current flat rate pricing. It is worth noting that although the user's price became predicable, the amount to be paid would vary in different periods as the coinsurance function to be offered would change according to the user's usage and behavior.

No frequent decisions required

Moreover, this insurance mechanism does not require frequent decision making. The amount of insurance to purchase would be decided only once in a period, say one month. In fact, the ISPs can always change the period time frame to better suit the users. To simplify the decision-making process, users could specify their desirable coverage rates to the ISP, allowing the ISP to automatically announce the price stability insurance payment in the beginning of the subscription period.

Flexible for users to choose their protection level

Another advantage of applying insurance is that this scheme is entirely flexible. Unlike conventional Internet pricing mechanisms that give users little choice or freedom, the insurance proposed here would not impose any obligation on users. Depending on the users' choice, users had the option to choose from purchasing no insurance to enjoying flat-rate pricing. In another sense, the introduction of insurance would provide users the freedom to tailor-made their best level of protection.

This chapter will be organized as follows: In Section 3.2, we mention the model used in this analysis, together with the Internet pricing process in detail. We then introduce the insurance setup in Section 3.3, where we mention the criteria to formulate the coinsurance function, and how users with different risk adversities response to such setting. In Section 3.4, we analyze the drawback of allowing users to subscript full insurance as the optimal solution, and introduce an alternate coinsurance function to solve problem.

3.2 The Internet Pricing Model

3.2.1 System Model

We consider a simple network model with a set $\mathcal{N} = \{1, \dots, N\}$ of users subscribing to an ISP. The ISP is connected to the rest of the Internet by a single egress link with capacity Y. For the economic model, we consider a third party insurance company is introduced. Users can choose to purchase insurance with the insurance company freely. An illustration of the system model is provided in Figure 3.1.

Each user n is characterized by a utility function $U_n(\cdot)$, which is nonnegative, increasing, and strictly concave. This is the typical assumption for elastic data applications.

We consider a discrete time model where a time period T is divided into several time slots t. For each time slot $t \in T$, we assume the ISP observes the network congestion situation and announces the congestion unit price price $p(t)$ to all users. The congestion unit price can be observed by recording packet loss situation in the network or by collecting ECN marks in the TCP

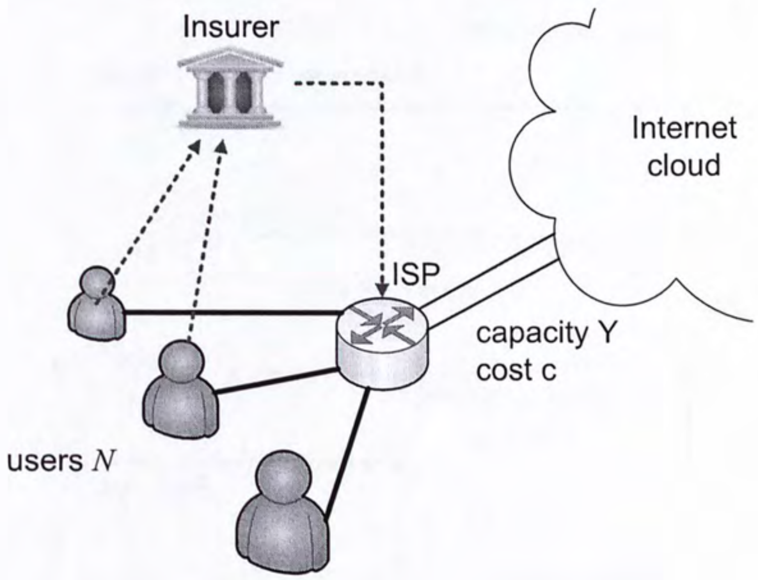

Figure 3.1: The system model adopted in this analysis.

header of packets. In contract and balancing mechanism, the congestion unit price reflects how much per unit traffic a user needs to pay when his traffic exceeds that contracted. Given user n's utility function, we can derive its demand as a function of the congestion unit price $p(t)$,

$$D_n(p(t)) = \arg\max_{0 \le x_n \le x_n^{\max}} \left(U_n(x_n(t)) - p(t)x_n(t)\right),$$

where $x_n \in (0, x_n^{\max})$ is the date rate that user n requests.

Given the value of $p(t)$, the demand function determines the optimal data rate $D_n(p(t))$ that maximizes the difference between an user's utility and the correspondent traffic cost. Since the utility function $U(\cdot)$ is concave, the demand function $D(\cdot)$ is convex and decreasing.

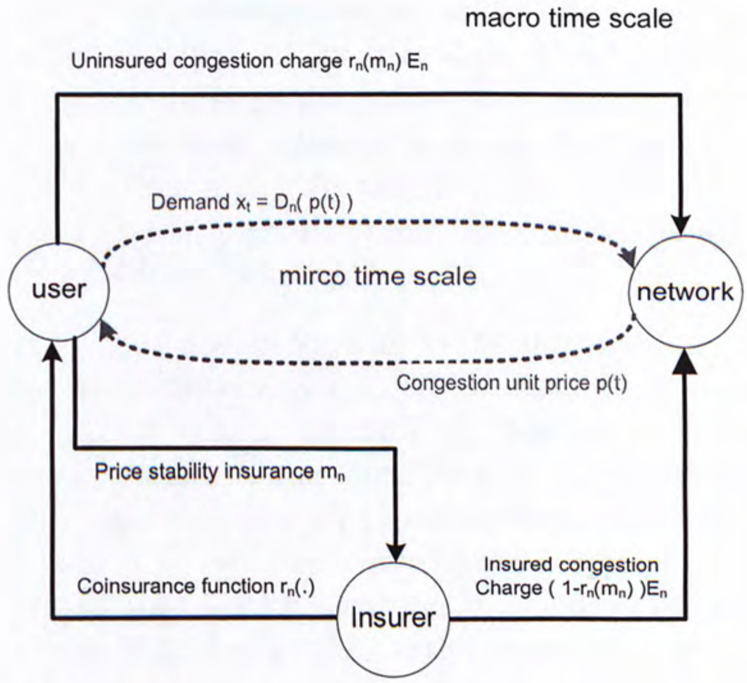

Figure 3.2: An overview of the charging scheme adopted in the charging model. Solid lines represented macro time events, while dotted lines represented micro time events.

3.2.2 Decisions Time Scales

Within each period T in our discrete time model, there are $t \in T$ time slots. Each time slot t is short enough that it is not possible for users to make any judgement or decision. An example of such duration would be half a second or one round trip time (RTT). The duration of T, on the other hand, is long enough such that users are comfortable to make long-term decisions. Such duration may be one month or more.

We divide the time period into two decision time scales to better describe the decision behavior of users. An illustration of the two time scales is provided in Figure 3.2.

1. The *micro time scale*: This time scale defines how users

fine-tune their immediate use of the Internet to maximize their utility surplus. The time scale of such is a time slot t. As fine-tuning behaviors in this time scale is not welcomed by users, we want to minimize the effect of decision making in micro time scale. At the same time, we want users to be cautious about their traffic, in order to provide higher network efficiency to the Internet.

2. The *macro time scale*: This is the time scale for users to make strategic planning to achieve maximum utility. The time scale of such is a period T. In this time scale, users find it comfortable to decide their long-term Internet usage and to fine-tune their cost to maximize their surplus from using the Internet. We want users to make critical decisions in this time scale, and become less conscious of pricing details at micro time scales. That is the reason why the insurance purchasing process take place in this time scale.

3.2.3 Micro Time Scale Pricing

We assume the pricing in micro time scale is performed by contract and balancing system. We consider the case when no prepayment is made for users, this is to allocate a fixed capacity charge among different users based on the proportion of congestion volume each user incurs. For each micro time t, the ISP announces the congestion unit price $p(t)$ to all users, and the users respond with their respective data rate $x(t)$ by the demand function $D_n(p(t))$. The congestion volume incurred by the users are therefore $x(t)p(t)$. At the end of the macro time, the payment for each user is calculated by the proportion of congestion volume compared to the total congestion volume of all users. The payment of user n at the end of the macro time is:

Definition 3.2.1. *The resultant payment of user n at the end of a period T under contract and balancing mechanism is defined as:*

$$E_n = \frac{\sum_{t \in T} x_n(t)p(t)}{\sum_{t \in T} X(t)p(t)} cY,$$

where $X(t)$ is the total traffic of all users at time t, Y is the network capacity and c is the per-unit cost of capacity announced by the ISP. For example, if the total congestion volume of a user contribute 10% of the total congestion volume, he is expected to pay $\frac{cY}{10}$ to the ISP at the end of the period. It is important to note that the total payment received by the ISP in such case is always cY. That does not only remove the traffic-cheating incentive, but also allows ISP to obtain return on investment at all times.

3.2.4 Macro Time Scale Pricing

The macro time scale pricing is the insurance purchasing process. In the beginning of a period, the third-party insurer announces a coinsurance plan to each network user. The coinsurance plan is a tariff function, indicating how much insurance payment would lead to what amount of protection. We name this coinsurance plan a *coinsurance function* for future reference. Note that the coinsurance function is tailor-made for each user, depending on the user's previous statistical behavior on Internet usage and traffic pattern. This is similar to health insurance, where the insurer customizes users' insurance plan individually based on the individual's medical background.

In this analysis, we assume the insurance company is able to obtain certain pricing information, such as a user's congestion charges history. We believe this information should be easy to retrieve. One way of achieving this is to ask users to submit their payment invoices to the insurer for each period.

After the coinsurance function is known, each user then chooses how much insurance one would like to purchase, based on the demand on network usage and the degree of risk adversities towards price/data rate fluctuations. Upon purchasing insurance, a portion of one's payment due to congestion can be alleviated. User n's overall payment for the entire macro time T is calculated as:

Definition 3.2.2. *The total charges a user n pays in macro time period T is*

$$m_n + r_n(m_n)E_n,$$

where m_n is the price stability insurance payment and $r_n(\cdot)$ is the coinsurance function. E_n is the payment from the contract and balancing mechanism in micro time scale from Definition 3.2.1.

With the effect of the m_n in the coinsurance function, the second term of the equation is small when enough price stability insurance is purchased. Therefore, serious price fluctuations from micro time transactions can be alleviated. A list of parameters used in this analysis is presented in Table 3.1.

We will present how the price stability insurance is set up in the next section.

3.3 Actuarially Fair Coinsurance Function

In this section, we investigate how the insurer design the coinsurance function $r(\cdot)$ and how users respond to the coinsurance function by purchasing price stability insurance.

We then introduce some of the concerns that may affect the users' choice of insurance payment. For instance, how this insurance payment affects the behavior of users, and how asymmetric information changes the nature of insurance. Based on the problems, we fine-tune our original coinsurance function mechanism

Table 3.1: Parameters used in the model

System parameters	
$\mathcal{N}$	A set of users subscribed to the ISP
$U_n(\cdot)$	Utility function of user n operating application A
$D_n(\cdot)$	Demand function of user n operating application A
Y	network capacity of the ISP
c	cost per unit capacity
Pricing parameters	
$p(t)$	Congestion unit price at time slot $t \in T$
E_n	Congestion charge of user n for the current period
ρ	Congestion probability
Insurance parameters	
$r_n(\cdot)$	Actuarially fair coinsurance function for user n
$r_n^p(\cdot)$	Premium coinsurance function for user n
m_n	Insurance purchased by user n
β	Base congestion charge in the coinsurance function
α	Premium factor

to remedy these problems in Section 3.4.

3.3.1 The Actuarially Fair Coinsurance Function

The first question we want to ask is to determine the most appropriate coinsurance function for different users $r(\cdot)$. One way to setup the coinsurance function is through calculating the actuarially fair coinsurance function. The actuarially fair coinsurance function is to ensure the insurer does not gain any expected profit by issuing the insurance, that is the insurance company should have zero expected profits, as defined in Definition 3.3.1.

Definition 3.3.1. *The actuarially fair coinsurance function $r(\cdot)$ defined such that the following condition is true:*

$$\rho(m - I(r)) + (1 - \rho)(m) = 0,$$

where $I(.)$ is the indemnification function by the insurer. The indemnification function is defined as the amount of money the insurer needs to pay when the insurant execute their insurance. In this scenario, the indemnification is the coinsured part of the user's contract and balancing mechanism payment the insurer needs to pay. Moreover, we use the variable $\rho \in (0, 1)$ to indicate the probability of congestion.

Definition 3.3.1 is reasonable even if the third party insurance company is not a non-profit organization. In order to better illustrate this issue, let us classify users into different risk adversity levels and explain how they react to risk.

We classify users into three types according to their different reactions to risk. These are risk-seeking users, risk-neutral users and risk-averse users. The behavior of these three types of users are shown in Figure 3.3. Figure 3.3 is referred as the expected utility function, or von Neumann-Morgenstern utility function in economics [41]. The term x can be any entity, one example of such is the amount of money acquired.

Risk-seeking users (asterisk line) are presented as a convex curve. Their utilities of the expected value is less than their expected utility from uncertainties. This is due to the fact that they gain higher satisfaction when they obtain higher payoff, even though that possibility is small. As an illustration, when such users play a coin-flipping bet, where they get one upon winning and get nothing when they lose (the expected value is 0.5), they are willing to pay higher than 0.5 for the bet.

Risk-neutral users (asterisk line) are presented as a straight line. Their utilities of the expected value equals to their expected utility from uncertainties. That means their utility is proportional to the quantity of entity they obtain. In the coin-flipping game, they are willing to pay up to 0.5 to play.

Risk-averse users (triangle line) are presented as a concave curve. Their utilities of the expected value is greater than their expected utility from uncertainties. For the coin-flipping game with expect value of 0.5, they are only willing to pay less than 0.5 for the game.

As an insurance company, it could make profit by bridging the utility difference between a risk-averse user and the expected revenue obtained (that is to absorb the risk premium of the user). Since risk-averse users tend to pay more than their expected utility, the insurance company can make a profit by this difference in exchange for a risk-protection insurance plan.

By calculating Definition 3.3.1, we are able to obtain the actuarially fair coinsurance function in this model:

Theorem 3.3.2. *The actuarially fair coinsurance function $r_n(\cdot)$ of user n under contract and balancing mechanism defined in Definition 3.2.1 is:*

$$r_n(m_n) = 1 - \frac{m_n}{\rho E_n},$$

The term ρE_n is the expected payment under contract and balancing mechanism for the macro time T. We refer this term

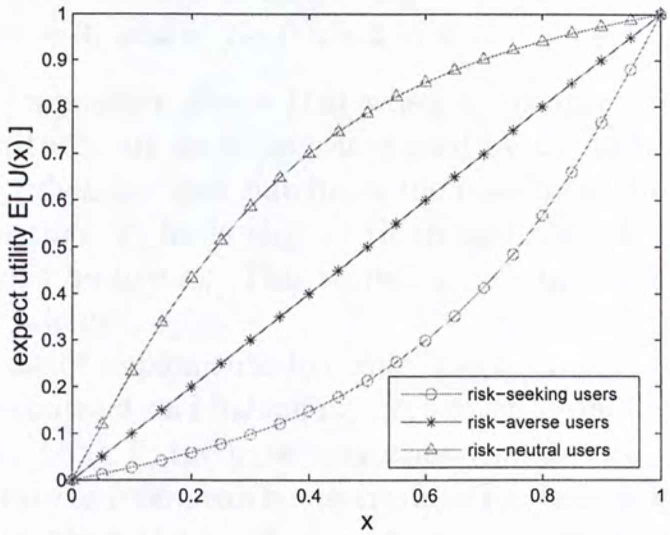

Figure 3.3: The expected utility function for users with different risk adversities.

as the base congestion charge and represent it as β_n in later context. Based on the traffic behavior of different users, the actuarially fair coinsurance function are different. In the beginning for each macro time, the ISP announces this coinsurance function to different users with different *base congestion charges*. Then the users respond with their optimal insurance charges. Note that the smaller the value of $r(\cdot)$, the greater the user is being insured. The proof of Theorem 3.3.2 is stated in Appendix A.1.1.

3.3.2 Properties of the Actuarially Fair Coinsurance Function

After the actuarially fair coinsurance function is defined, we analyze the properties of the coinsurance function for different types of users.

Proposition 3.3.3. *The actuarially coinsurance function* $r(\cdot) = 1$ *when* $m = 0$, *and* $r(\cdot) = 0$ *when* $m = \beta$.

This proposition shows that when no insurance is purchsed ($m = 0$), there are no coinsurance paid by the insurer ($r = 1$); however, when the user purchases the base congestion charge β in macro time T, he is eligible to enjoy 100% discount for all congestions he incurs. This turns the pricing mechanism into flat rate pricing.

In terms of implementation, since the actual congestion payment by contract and balancing mechanism is only reveled after the macro time T, the base congestion charge used in the coinsurance rate function can be determined by statistical average of a user's traffic pattern and congestion incurred. We will discuss the implementation details in our numerical example in Section 3.3.4.

Proposition 3.3.4. *The actuarially fair coinsurance function* r *is monotonic decreasing with the price stability insurance* m.

This proposition shows that users always get better insured when they purchase more price stability insurance m. However, the actual protection obtained by each user is different for the same amount of payment, as the base congestion charge of different users varies, depending on individual's coinsurance function $r_n(\cdot)$.

3.3.3 How Much Insurance Should a User Buy?

After the actuarially coinsurance function $r(\cdot)$ is determined, one may ask how much insurance a user should purchase when such function is known? To answer this question, we need to solve the following optimization problem:

$$\max_m \rho V(w - m - \beta) + (1 - \rho)V(w - m), \tag{3.1}$$

Here we assume the congestion probabilities $\bar{\rho}$ are the same for the entire macro time T, that is $(\rho(t) = \rho, \forall t \in T)$ for simplification. Note that this only assumes the probability of congestion to be the same, however the degree of congestion $p(t)$ can vary for each time slot $t \in T$. Using Equation (3.1), we first find the optimal insurance charge m^* for risk-averse users:

Proposition 3.3.5. *The solution of Equation (3.1) for risk-averse users is:*

$$m^* = \beta$$

The proof of Proposition 3.3.5 can be obtained in Appendix A.1.2. This value equals to the base congestion charge. When such price stability insurance is paid, the portion of coinsurance paid by the user is:

$$1 - \frac{\beta}{\beta} = 0,$$

which a risk-averse user's best response of obtaining the coinsurance function is to purchase full insurance. This is equivalent to paying the total expected congestion charge in the beginning of the macro time, in order to prevent being charged extra in the coming macro time. Note that all risk-averse users of any degree purchase full insurance as their best responses because of the concavity of the expected utility of risk-averse users.

Proposition 3.3.6. *The solution of Equation (3.1) for risk-seeking users is $m^* = 0$.*

The proof of Proposition 3.3.6 can be found in Appendix A.1.3. The result is intuitive as risk-seeking users does not treasure stable payments. Their utilities raise when there are chances that their payments are smaller than the expected payment for some macro time. As a result, they tend not to purchase any insurance.

Proposition 3.3.7. *There are infinitely many solutions of Equation (3.1) for risk-neutral users in $0 \leq m^* \leq \beta$.*

The proof of Proposition 3.3.6 can be found in Appendix A.1.4. Risk-neutral users obtain the same utility for any m^*, as there is no difference between their utilities and the expected charges under this actuarially fair coinsurance function. They get the same utility surplus by buying (or not buying) any insurance.

3.3.4 Numerical Examples

We provide a realistic numeric example to show how the mechanism of insurance works in practice and how price stability insurance helps users to avoid experiencing huge price fluctuations.

In this example, we illustrate the effect of price stability insurance by a simple model. Before we describe the model in detail, let us define the utility function and demand function used in the model:

Definition 3.3.8. *The utility function of a user n transmitting at a data rate x is:*

$$U_n(x) = u_n \log(x), \tag{3.2}$$

where u_n is the magnitude of user n's utility. The greater the u_n, the more utility the user obtains by transmitting traffic. It can be used as a measurement to determine how important ones traffic is. Based on the above, we can formulate the demand of the user by solving the optimization problem stated in Equation 3.3.4:

$$D_n(p) = \frac{u_n}{p}.$$

The proof is intuitive and it is omitted here.

In the model, there are a random number of users subscribed to the network. We assume all users are risk-averse by nature.

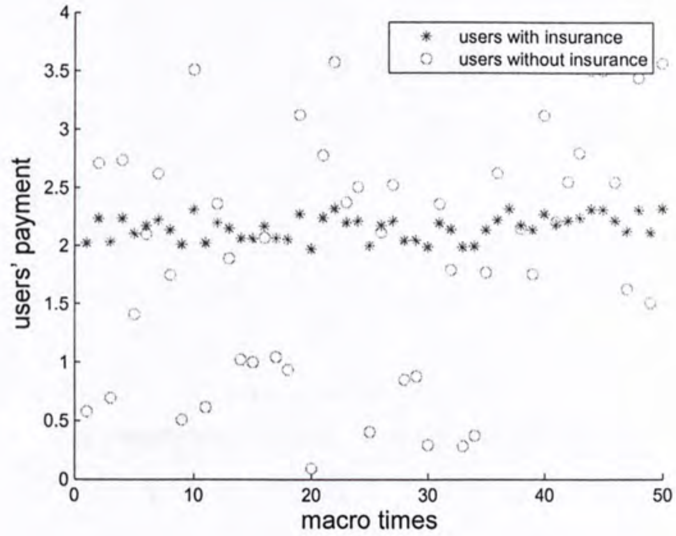

Figure 3.4: Payments for insured users (asterisk link) and uninsured users (circle line).

Each user n has a utility function with a distinct value u_n. We assume that, because there are so many users in the network, an user's alternation of data rate has virtually zero effect to the congestion of the network. We model the congestion unit price p as a random binomial distribution. We also assume that all users demand network traffic at all times.

In the beginning of each macro time scale, the insurance company announces the actuarially fair coinsurance function to each users. The base congestion charge the insurer sets to each user is the running mean average congestion charge of the previous ten macro times. In micro time scale, the ISP announces the congestion unit price $p(t)$ at each time slot t, and all users respond to the unit price with their demand function $D(p(t))$. At the end of the macro time, users settle the congestion payment by paying the uninsured portion of the congestion charge to the ISP based on the contract.

Figure 3.4 shows the users' payment across macro times. Al-

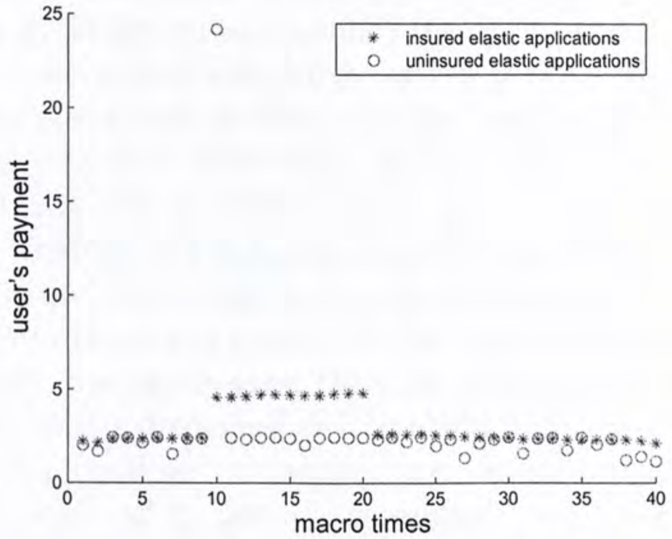

Figure 3.5: Payments for user with/without insurance charge when a sudden change of demand happens at macro time 10.

though the actual payment varies across macro times, users' payments are predicable at the beginning of the macro time if they have purchased full insurance.

Compared with users that have not purchased any insurance (circle line), purchasing price stability insurance help reducing the fluctuation of payment across macro times. Note that as actuarially fair rate is adopted in this example, the mean value of the two insurance total payments are the same. In such case, the best response of a risk-averse user is to purchase full insurance.

The coinsurance function is calculated based on the assumption that a user's behavior would not change upon purchasing the insurance. Next we look into the phenomenon when a user tries to deviate from one's normal behavior after one purchases some price stability insurance. Figure 3.5 shows a scenario where a user suddenly multiplies his traffic by 10 times at macro time 10. Assume the user with insurance (asterisk line), his payment at macro time 10 is far smaller than one without insurance (circle line) as a portion of the congestion charge is cushioned by the insurer. However, as a penalty, the user would suffer from a higher base congestion charge for ten periods when running mean average is used to determine the base congestion charge. Therefore, the best traffic response of a user in long run is to maintain ones normal traffic pattern.

As an analogy, the running mean average calculation is similar to the no-claim discount in auto-mobile insurance. An accumulative discount is applied to the insurance charges of the auto-mobile insurance when the user does not file any claims. When the user experiences any accidents that results in insurance claims (analogy to causing unexpected congestion), the discount wears off (a penalty is applied) and the user would be uninsured for a number of periods before enjoying the same discount rate again.

3.4 Premium Coinsurance Function

In the previous section, we showed that if the coinsurance function is actuarially fair, the best response of risk-averse users is to purchase full insurance, that is to pay the base congestion charge in advance. In the network's point of view, however, full insurance of users is not a preferred option for ISPs, as providing full insurance might change the behavior of users. In the insurers' point of view, it is difficult for the insurer to determine the full insurance coinsurance function by statistical information.

In this section, we show the problems raised by actuarially fair coinsurance function, and we propose an alternate coinsurance function to remedy the problems. We then analyze the responses of different users with a numerical example to better illustrate the performance of such insurance.

3.4.1 Problems of Allowing Full Insurance

The first problem of issuing full insurance is moral hazard. Moral hazard is the phenomenon that a user behaves differently when part of one's risk is insulated by insurance than under full risk. More information on moral hazard in economics can be found in [35, 39]. Without insurance, the demand of a user at micro time t is $D(p(t))$. However, when the user purchases some insurance $m \geq 0$ and enjoys a coinsurance of $r(m)$, there are incentives for the user to change his demand to $D(r(m)p(t)) > D(p(t))$ in the micro time. This phenomenon becomes more apparent when more insurance is bought. To an extreme, when full insurance is purchased, we can expected that the user's demand is always $D(0) = x^{\max}$.

Proposition 3.4.1. *The maximum deviation of demand happens when full insurance is purchased. Considering the following optimization problem:*

$$\max_{m} D\Big(r(m)p(t)\Big) - D\Big(r(m + \triangle m)p(t)\Big),$$

The solution of the optimization problem is

$$m = \beta,$$

that is full insurance as indicated in Proposition 3.3.5.

The proof can be found in Appendix A.1.5. One example of this phenomenon is dinning in a buffet. Since ordering extra in a buffet does not raise the total bill (ie. it is risk-free), people generally behave differently under such setting and consume more than under normal circumstance. If full insurance is the best response of the users, we can expected that users tend to over-consume network bandwidth and this leads to network inefficiency as we observe from the current flat rate pricing in the Internet.

The second problem of issuing full insurance is that the actuarially fair coinsurance function r is calculated using statistical information such as the past congestion charges of each user. As those records are totally statistical, the insurance company is at risk that the user will not behave similarly in the next macro time.

Assuming that the congestion unit price $p(t)$ in the actuarially fair coinsurance function in Theorem 3.3.2 is known before the macro time. For example, we can foresee that the congestion unit price would be larger during working hours, and the base congestion charge of the coinsurance function is determined by the user's traffic at the previous macro time. Now, the user who knows his upcoming usage in the next macro time purchases his optimal price stability insurance charge by solving the following optimization problem:

$$\begin{aligned} & \min_m \ m + r(m)E \\ \Rightarrow \ & \min_m \ E + m\Big(1 - \frac{E}{\beta}\Big), \end{aligned}$$

where β is the base congestion charges of the user, which is a statistical value of the previous macro time T_p and E is the congestion charge at the current macro time. The range of m is in $(0, a)$ and E is a function of $x(t)$, which is the known data rate of the user in the coming macro time. The solution of this linear function depends on the slope of the function:

$$m^* = \begin{cases} \beta & \text{when} \quad E_n \geq \beta, \\ 0 & \text{when} \quad E_n < \beta, \end{cases}$$

It shows that when the user acquires more information than the ISP does, the user will only choose between paying a full insurance or not paying any insurance at all. When the user has a better knowledge on the traffic pattern he will use in the next macro time $x(t)$ for example, he has a better idea on the next congestion charge E_n, and can optimize his choice of insurance using the above rationale. This causes the insurance mechanism to fail as the insurance is implemented to alleviate price fluctuations, rather than acting as a tool to arbitrage user's utility surplus.

To remedy the problem of issuing full insurance, we need to adjust the actuarially fair coinsurance function such that risk-averse users' best responses are not paying full insurance. One method to achieve this is to set up another coinsurance function that apply some extra charges if users require full insurance, as explained in the next subsection.

3.4.2 The Premium Coinsurance Function

In this subsection, we formulate the premium coinsurance function to achieve the following two requirements:

1. Full insurance should not be the best response of users. Risk averse users should not purchase full insurance for their best response in order to mitigate the data rate deviation caused by moral hazard, and;
2. A premium in addition to the base congestion charge is needed for users to purchase full insurance, so that it is still open for users to purchase even though full insurance is not an optimal response.

One way to achieve the above two criteria is to implement a premium factor α to the actuarially fair coinsurance function $r(\cdot)$ as follows:

Definition 3.4.2. *The premium coinsurance function rate for user n is defined as:*

$$r_n^p(m_n) = 1 - \frac{m\alpha}{\rho E_n},$$

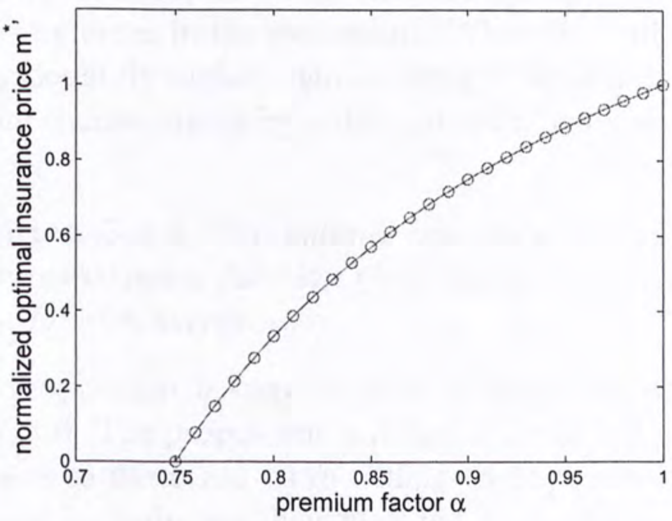

Figure 3.6: When premium factor α increases, less premium is added to the actuarially coinsurance function, therefore the optimal insurance payment of users increase. Here we normalized the base congestion charge as 1.

where $\alpha \in (0, 1)$. The premium coinsurance function equals to the actuarially fair coinsurance function when $\alpha = 1$. We first analyze this premium rate r^p by determining the optimal insurance charges of users with different risk adversities:

3.4.3 Properties of the premium coinsurance function

Proposition 3.4.3. *The optimal price stability insurance m^{p*} under premium coinsurance function r^p is increasing in α for risk-averse users.*

The proof of the proposition is shown in Appendix A.1.6. Figure 3.4.3 shows one example of the relationship between the premium rate distortion factor α and the optimal insurance charge m^{p*}. We can see that as the premium factor α decreases (more premium is required to obtain the same coinsurance), the optimal insurance payment m^{p*} decreases as well. This is similar to the dinning scenario, where one needs to decide between a buffet or charge by order in the restaurant. When the buffet charge is unproportionately higher than ordering it separately(α is low), one would choose charge by order and not a buffet despite being risk-averse.

Proposition 3.4.4. *The optimal insurance charge m^{p*} under premium coinsurance function r^p is increasing in users' initial weight w for risk-averse users.*

This proposition is easy to observe from the result of Appendix A.1.6. The proposition is intuitive to the fact that wealthier risk-averse users are more willing to pay extra for greater satisfaction by reducing their payment fluctuations. Note that in premium coinsurance function, the more insurance one buys, the greater the mean total payment one needs to pay. This is different from the actuarially fair coinsurance function that the mean total payment is independent to the amount of insurance purchased as shown in the numerical example in Figure 3.4.

Proposition 3.4.5. *The optimal insurance payment m^{p*} of risk-seeking and risk-neutral users under premium coinsurance function r^p is $m^{p*} = 0$.*

The proof of the proposition can be found in Appendix A.1.7. The proposition follows that since the premium coinsurance function is unfair to users(in a sense that users need to pay more in average when they purchase the same discount rate), users who are not averse to risk should not buy this insurance.

In general, users are generally risk averse in nature [19]. Therefore we can expected that there are demand for premium coinsurance function when it is released, albeit its "unfair" nature. Note that full insurance is still possible if users are willing pay $\frac{\beta}{\alpha}$. This price, however, is not a best response to users when $\alpha < 1$.

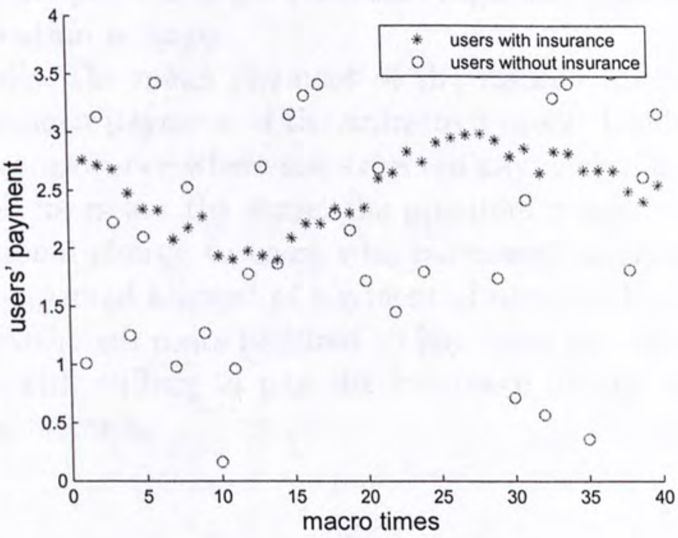

Figure 3.7: Payments for insured users (asterisk link) and uninsured users (circle line) under premium coinsurance function $r^p(\cdot)$.

3.4.4 Numerical Example

In this subsection, we want to use an numerical example to illustrate how premium coinsurance function works under a realistic model.

The environment setting of this model is the same as the numerical example in Section 3.3.4, except that the coinsurance function used in this example is the premium coinsurance function. We set the premium factor α to 0.7 for the simulation.

Figure 3.4.4 shows the result of the simulation. There are two observations in the simulation. Firstly, similar to the actuarially fair coinsurance, the premium insurance can help reducing the fluctuation of users who bought insurance, however, we can see the range of variance is larger than in actuarially fair coinsurance. It is because under premium coinsurance, the best response of user is to purchase less insurance than the full insurance. With less insurance purchased, the user is required to bear a larger portion of the eventual congestion charge and thus the fluctuation is larger.

Secondly, the mean payment of the insured users is higher than the mean payment of the uninsured users. Unlike actuarially fair coinsurance where the expected payment of insured and uninsured users are the same, the premium coinsurance incurs an additional charge to users who purchased insurance, therefore the expected amount of payment of users with insurance is higher. Although users required to pay more for the insurance, they are still willing to pay the insurance for an appropriate premium factor α.

□ **End of chapter.**

Chapter 4

Problems in Revenue Distributing Stage

Summary

The Internet includes thousands of Internet service providers (ISPs), which are interconnected to provide connectivity and service for end users. [illegible] the settlement between the ISPs are determined based on bilateral agreements that result from pairwise negotiations. Although this settlement mechanism is [illegible] and easy to implement, it does not encourage network-wide cooperation [illegible] the Internet. [illegible] not lead to a fair division of revenue among all ISPs that are involved in carrying the traffic [illegible]. This problem is getting more severe with [illegible] new Internet business models.

In this paper we try to [illegible] [illegible]

Chapter 4

Problems in Revenue Distributing Stage

Summary

The Internet includes thousands of Internet service providers (ISPs) which are interconnected to provide connectivity and service for end-users. Traditionally, the settlement between the ISPs are determined based on bilateral agreements that result from pair-wise negotiations. Although this settlement mechanism is intuitive and easy to implement, it does not encourage network-wide cooperation, as the bilateral charges typically do not lead to a fair division of revenue among all ISPs that are involved in carrying the same flows of traffic. This problem is getting more severe with various emerging new Internet business models.

In this paper, we try to determine the existence and realizability of bilateral prices that can achieve fair revenue division among ISPs. In particular, we use Shapley value as the basis for deriving fair prices. Under a quite general topology and traffic model, we find that there exists prices that make the revenue division under bilateral settlement equal to that calculated under Shapley value. The corresponding "fair price" exhibits several nice and desirable characteristics. Moreover, it could be realized approximately.

4.1 Introduction

The Internet is the result of interconnecting many networks, each operated by a separate ISP. The Internet service, as enjoyed by the customer of all ISP networks, is however a service provided collectively by all ISPs. A customer of the Internet service subscribes to an access ISP. For this service, the access ISP collects a charge from each subscriber. Since the service is accomplished by multiple ISPs, a fundamental issue is how ISPs should divide up this charge.

Based on the design of the Internet, the unit of network service is a packet transported. This is a rather miniscule unit for measuring service, not to mention the task of dividing the charge among multiple parties contributing to the service. By convention (established historically), ISPs charge users monthly on a flat rate basis (like an all-you-can-eat buffet), and settle account monthly on a bilateral basis between ISPs who are connected with each other. There are two most common types of bilateral peering relationships between ISPs. In the first type of peering relationship, one ISP is considered as a transit provider for the other ISP, and the transit provider charges its customer ISP for amount of traffic transit through the provider network. The second type of peering relationship is a totally collaborative one. Two ISPs exchange traffic and deem to benefit from it mutually and forego any charges to each other. Although this all seems a rather sloppy business practice, it keeps the effort in book-keeping to a minimum. This minimalist approach is also totally decentralized, without the need for all kinds of coordination between ISPs.

In recent years, new business models (using Internet services) emerge; for example, Internet content providers (ICPs) are able to generate significantly higher levels of revenues [31]. This prompted re-examination of how the benefits of network services should be divided between various players. One study [28] specifically re-examined the issue of how ISPs should share the total revenues of the Internet service and proposed Shapley values as a potential solution. Shapley value refers to an axiomatically derived formula for fairly dividing a prize among a set of contributors in a general economic setting. By forming an coalition, it is argued in [28] that the ISPs will find more optimal routes and maximize the overall value of the Internet.

In this work, we ask a different question: Can the current bilateral peering between ISPs, under proper pricing strategies, lead to a cooperative settlement as if the ISPs are all in a coalition? This proposition is not entirely unreasonable, since ISPs do realize the positive network externality in the interconnected network. There are actually two parts to this question: a) does there exist fair pricing (defined as prices in bilateral peering that produce Shapley value as settlements)? and b) how feasible and likely ISPs will choose such fair prices in their peering agreements. For a general set of ISP network topologies and traffic models, we show the answer to (a) is true. We then show some preliminary results to (b).

This work is organized as follows. We formulate the system model and settlement model in section 4.2 and 4.3. In section 4.4, we show that in a symmetric network setting, we could obtain the fair prices that produce Shapley value and show some desirable properties of the fair prices in Section 4.5. We further consider how fair prices can be calculated in the more general asymmetric topology in Section 4.6. In Section 4.7, we show how the fair prices can be locally approximated without knowing the global network information.

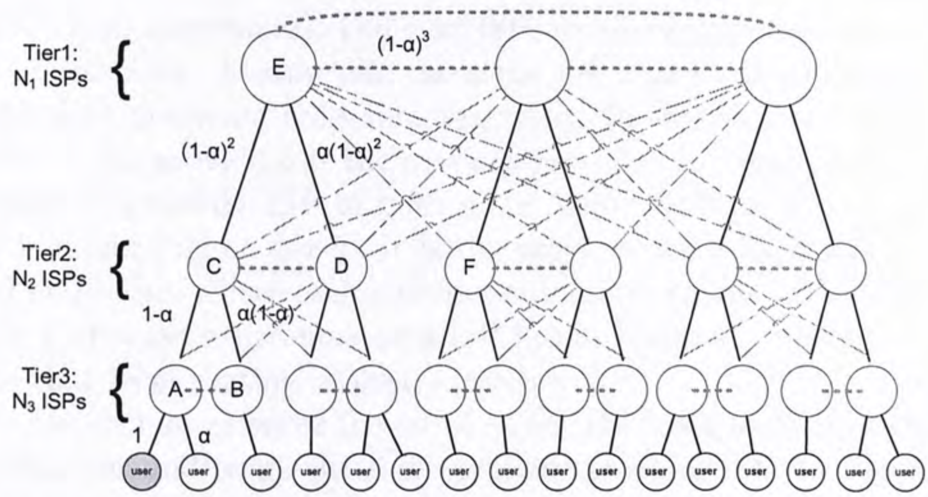

Figure 4.1: An illustrative example of the topology and traffic model. Here we have three tiers of ISPs, with three different links: transit links (solid lines), peering links (dotted lines), and multi-homing links (dash lines). For a unit of traffic sent by an end-user of leaf ISP node A (greyed circle), the values beside some of the links represent the probability that the traffic goes to a particular height.

4.2 System Models

4.2.1 Topology Model

We use an AS-level hierarchical model to capture the essence of the current tier-based Internet. The hierarchical model has been extensively used in related literatures (eg. [20], [21]). We assume that each ISP contains only one AS.[1] Figure 4.1 is an example of such topology. A set of $\boldsymbol{\mathcal{N}} = \{\mathcal{N}_t, \forall t \in \{1, \ldots, T\}\}$ ISPs is located at different tiers, where the set of ISPs in tier t is denoted by $\mathcal{N}_t$. The highest tier (tier 1) is the backbone tier and the lowest tier (tier T, $T = 3$ in Figure 4.1) connects to end-users. For a particular ISP node $n \in \mathcal{N}_t$ with $t \in \{1, \ldots, T - 1\}$, it serves as the *provider* for a fixed number of tier $t + 1$

[1] For an ISP that contains more than one AS, it is enough to think it as one AS in our model.

ISPs (i.e., *customers*). Different ISPs can have different number of customers. Meanwhile, the same ISP can purchase several different upstream links simultaneously. In this case, we define one of the providers as the *primary provider*. In Figure 4.1, the primary provider ISP of both node A and node B is node C. For a leaf ISP at tier T, it is the provider for a fixed number of end-users, where each end-user subscribes to one tier T ISP. Furthermore, we assume each ISP has at least two subscribers, so that local routing is always feasible.

We define *region* as the set of nodes which are located at the same tier and share the same primary provider ISP. In Figure 4.1, node A and node B are in the same region; node D and node F, node A and node C are in different regions.

We classify the links into three types:

1. *Transit link*: a cross-tier link that connects an ISP node to its primary provider ISP at the upper tier. All ISP nodes in the same region establish a transit link to the same provider ISP. In Figure 4.1, the solid lines represent transit links. Node C and node D are both connected to their common primary provider ISP node E through different transit links.

2. *Peering link*: a link that connects ISPs in the same tier within the same region. Peering links are marked as dotted lines in Figure 4.1. There is a peering link between node C and node D, but there can not be any peering link between node D and node F since they are not in the same region.

3. *Multi-homing link*: a link that allows an ISP to connect to another upper-tier ISP which is different from but in the same region as its primary provider ISP. In Figure 4.1, multi-homing links are marked as dash lines. Node A establishes a multi-homing link to node D, but can not establish a multi-homing link to node F since it is not in the same region as its primary provider ISP node C.

Next we define several network topology parameters. First we formulate the peering strategy among ISPs in a macroscopic manner. For all nodes in the same region, we assume that there either does not exist any peering link, or there is a peering link between any pair of nodes such that they form a full mesh topology.[2] We denote $\boldsymbol{\rho} = \{\rho_n, \forall n \in \boldsymbol{\mathcal{N}}\}$ as the set of probabilities that such mesh connection of peering links exists at the region of each node n. Notice that if two nodes n and m are in the same region, then $\rho_n = \rho_m$. Since in practice the topmost tier (tier 1) nodes are typically interconnected into a full mesh through peering links ([20], [32]), we let $\rho_n = 1$ for all nodes $n \in \mathcal{N}_1$.

We denote the multi-homing degree $\boldsymbol{d} = \{d_n, \forall n \in \boldsymbol{\mathcal{N}}\}$ as the number of high-tier ISPs that each node n connects to. When $d_n = 1$, ISP n only connects to its primary provider ISP and thus there is no multi-homing. In Figure 4.1, $d_C = d_D = 3$ and $d_A = d_B = 2$. We assume that a node is able to use any of its multi-homing links to deliver traffic to all destinations.[3]

In later sections, we will denote a topology $\mathcal{G}$ as $(T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}, \boldsymbol{d})$. The list of parameters used in the paper can be found in Table 4.1.

4.2.2 Traffic Model

We adopt the traffic model used in [21]. In this model, only end-users (who subscribe to the leaf tier T ISPs) are able to initiate and terminate traffic (i.e., act as source and destination). To formulate this user-to-user traffic, we use a parameter α to illustrate how far the traffic is likely to go, instead of specifying the exact destination of the traffic.

[2]The "either none or full mesh" assumption is used to guarantee all traffic that need to go from one ISP node to another ISP node in the same region experience the same hop count: either two-hop by going through the primary provider ISP or one-hop by going through one of the peering links.

[3]This assumption is used to simplify the calculation of Shapley value. In reality, however, there may be some destinations that only some of the multi-homing links can reach.

For node $n \in \mathcal{N}$, there is a probability α_n for the traffic to go local with intensity I_n. For example in Figure 4.1, when an end-user of node A has traffic parameter α, his traffic terminates at another end-user of node A with probability α, goes to node C with probability $(1-\alpha)$, and terminates at an end-user of node B with probability $\alpha(1-\alpha)$. The furthest a traffic can go is with probability $(1-\alpha)^T$, which the traffic reaches an end-user that needed to transverse through a peering link at tier 1.

We assume BGP routing is used in the network. In BGP routing, ISPs connected with peering links only exchange routes of its customers to each other. As a result, traffic routes in the hierarchical model have hill-shape appearances (e.g., [14]) - they go uphill until the highest necessary tier, go flat through the peering link if needed, and go downhill to the destination. Because of this, the smallest number of tiers that a traffic reaches essentially determines the distance of that traffic.

4.3 Settlement Model and Definition of Fair Price

Next we describe the two settlement mechanisms to be studied in this paper, namely the bilateral peering settlement model (simplified as *Bilateral*) and the Shapley value cooperative settlement model (simplified as *Shapley*). Then we define a set of *fair prices* that lead to the same revenue division in both settlement mechanisms.

4.3.1 Bilateral Settlement

Bilateral is the settlement that is being used in the current Internet. In *Bilateral*, the charging between two ISPs are determined by mutual agreements [3].

Let us first consider the charging between end-user and leaf

Table 4.1: Parameters used in the model

Topology model	
T	Number of tiers in the model
$\boldsymbol{\mathcal{N}} = \{\mathcal{N}_t, \forall t \in \{1, \ldots, T\}\}$	Set of ISP nodes in the model
$\boldsymbol{\rho} = \{\rho_n, \forall n \in \boldsymbol{\mathcal{N}}\}$	Probability of existing peering links in mesh in the region
$\boldsymbol{d} = \{d_n, \forall n \in \boldsymbol{\mathcal{N}}\}$	Multi-homing factor - number of upstream connections
$\mathcal{G}$	A topology of parameters $(T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}, \boldsymbol{d})$
Traffic model	
$\alpha_n, n \in \boldsymbol{\mathcal{N}}$	Probability that the traffic goes local
$I_n, n \in \boldsymbol{\mathcal{N}}$	Intensity of traffic
(a, b)	A traffic from source a to destination b
Settlement model	
γ	Bilateral charing per unit traffic
$\varsigma_n,\ \psi_n(h, q, \boldsymbol{d}), n \in \boldsymbol{\mathcal{N}}$	Shapley value
$\mathcal{B}_n, n \in \boldsymbol{\mathcal{N}}$	Betweennes
$\boldsymbol{P}^* = \{P_t^*, t \in [1, \ldots, T-1]\}$	The fair price

tier T ISPs, which is the same in both settlement schemes. We assume that each end-user generates the same amount of traffic to the network, and is charged a fixed amount P_f by its provider ISP at tier T. The total amount of revenue collected from all end-users in the whole network is denoted as P_a.

For settlements between ISPs, we consider a per-traffic tier-based charging scheme. The precise charging scheme depends on the type of link between the two ISPs. Note that for simplicity we assume that the price is tier-dependent but not node-dependent. This is justified in a macroscopic view that ISPs in the same tier have similar size and thus have similar negotiation power in coming to a bilateral price. If the link is either a transit link or multi-homing link between tier t provider ISP and its tier $t+1$ customers, a charge P_t is paid by the customer to the provider for per-unit of traffic regardless of the traffic direction. If the link is a peering link between two ISPs, no charging is involved for any traffic sent over this link in either direction.

For the ease of later discussions, let us calculate the revenue obtained by an ISP n at tier t for handling (i.e., receiving and sending) one unit of traffic as the following:

$$\gamma_{n,t}(\boldsymbol{P}) = \begin{cases} 2P_t, & \text{for local traffic,} \\ P_t, & \text{for peering traffic,} \\ P_t - P_{t-1}, & \text{for upstream traffic,} \end{cases}$$

for $t \in [1, \ldots, T-1]$, and

$$\gamma_{n,T}(\boldsymbol{P}) = \begin{cases} -P_{T-1}, & \text{for upstream traffic,} \\ 0, & \text{for local/peering traffic,} \end{cases}$$

where $\boldsymbol{P} = \{P_t, \forall t \in \{1, \ldots, T-1\}\}$.

For example, in Figure 4.1, node C at tier 2 earns $2P_2$ when it handles a local traffic such as the one from A to B through C, earns P_2 when it handles peering traffic such as one from A to D, and earns $P_2 - P_1$ for upstream traffic such as A to E. For other traffic that does not pass through node C, it earns nothing. As leaf ISPs already obtain flat rate charging P_f from the end-users, they do not obtain any additional traffic-based revenue from the end-users. In Figure 4.1, node A loses no revenue when the traffic goes to end-users of node A and B(if there exists a peering link between A and B). It however needs to pay P_{T-1} when the traffic goes upstream.

In this settlement method, each ISP keeps a traffic counter at its ingress links to record the total amount of traffic handled at each link. At a specific clearance time, typically once a month, ISP t charges their customer ISPs P_t times the traffic passed through their transit/multi-homing links.

4.3.2 Shapley Settlement

Shapley is the settlement that divides the revenue among ISPs according to Shapley value. It is previously studied in [28], where the authors showed several nice properties of implementing such settlement in the Internet.

Shapley value uses an axiomatic approach to allocate benefits obtained from a coalition among all participating players. It satisfies the following four axioms:

1. *Efficiency*: players distribute among themselves the resource available to grand coalition.

2. *Symmetry*: if two players are symmetric, their payoffs are the same.

3. *Dummy*: dummy player receive no payoff.

4. *Additivity*: the Shapley value of a combined game is equal to the Shapley value of the separated games.

The Shapley value is proofed to be the unique solution that satisfying the above four axioms. Detailed information on Shapley value can be found in Shapley[38] and Winter [43].

Shapley is similar as *Bilateral* in terms of how the end-user is charged (i.e., flat-rate charging with P_f for each user and P_a for all users). The key difference is how different ISPs settle the charge among each other. Instead of charging differently for provider-customer relationship and peer-to-peer relationship, *Shapley* divides the revenue obtained from the end-users among ISPs by calculating the Shapley value of each ISP for each flow of traffic. The calculation of Shapley value is formally defined as follows:

Definition 4.3.1 (Shapley value). *The Shapley value that node $n \in \mathcal{N}$ obtains through handling traffic (a, b) (i.e., end-user a to end-user b) is:*

$$\varsigma_n(a,b) = \frac{1}{|\mathcal{N}|!} \sum_{\pi \in \mathcal{N}} [v(\mathcal{S} \cup \{n\}, a, b) - v(\mathcal{S}, a, b)],$$

where $|\mathcal{N}|$ is the total number of nodes in the network, π is the set of possible permutations of $\mathcal{N}$, $\mathcal{S}$ is the subset of ISP nodes in a permutation that appears earlier than n, and $v(\mathcal{S}, a, b)$ is the characteristic function of Shapley value for the traffic from a to b under the sub-topology formed by the set $\mathcal{S}$. We first note that Shapley value is always between 0 and 1. Furthermore, Shapley value calculates the contribution of a player (a node in our case) as the normalized marginal contribution under all possible ways of including it into a set of nodes $\mathcal{N}$. The measurement of worthiness is reflected by the characteristic function as defined next:

Definition 4.3.2 (Characteristic function of Shapley value). *The characteristics function $v(.)$ for a set of nodes $\mathcal{S}$ handling* *a traffic from a to b is:*

$$v(\mathcal{S}, a, b) = \begin{cases} 1 & , \textit{if there exists a path in } \mathcal{S} \textit{ that} \\ & \quad \textit{connects a to b,} \\ 0 & , \textit{otherwise.} \end{cases}$$

This characteristic function concerns whether the traffic can be completed with set $\mathcal{S}$. When there is already a complete route from source to destination ($v(\mathcal{S}, a, b) = 1$), all extra nodes that join $\mathcal{S}$ later have zero marginal contribution ($v(\mathcal{S}', a, b) - v(\mathcal{S}, a, b) = 0, \mathcal{S}' \supset \mathcal{S}$), as they are considered redundant in terms of providing connectivity. Two examples of Shapley value is shown at Figure 4.2a and 4.2b. When we consider a traffic from an end-user of node A to an end-user of node B. The connectivity breaks when either A or B is removed. However, the connectivity remains when either C or D is removed. Therefore, A and B have greater contribution than C and D, thus they obtain greater Shapley values.

To precisely implement the Shapley settlement, it is necessary to have a centralized network entity who calculates $\varsigma_n(a, b)$ for each ISP n for each traffic (a, b). At a specific clearance time, each ISP node uses the aggregated Shapley value calculated at the central entity to claim its portion of the total revenue P_a.

Based on the hierarchical model presented in this paper and BGP routing, it is possible to calculate the Shapley value based on a smaller set of parameters related to the traffic (a, b) as follows:

Definition 4.3.3. *The Shapley value of n can be reformulated as*

$$\psi_n(h, q, \boldsymbol{d}) = \varsigma_n(a, b),$$

where h is the height for traffic (a, b) (i.e. how far upstream the traffic goes), $q \in \{0, 1\}$ is the topmost peering parameter which states whether the traffic goes through a peering link at its

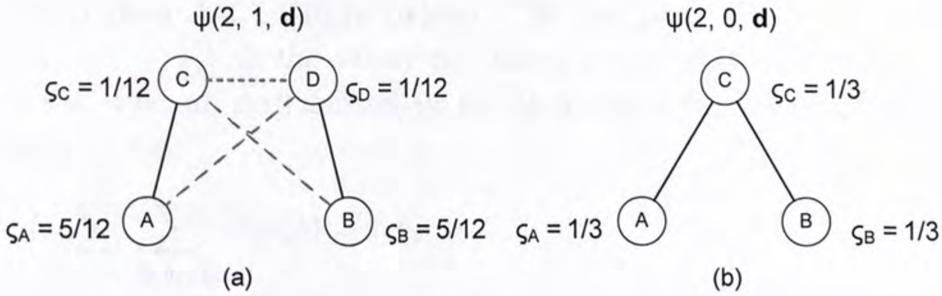

Figure 4.2: An example of how Shapley value is calculated. We consider a traffic from an end-user of node A to an end-user of node B, and calculate the Shapley value of each node with respect to such traffic. Here the multi-homing factors are set as $\boldsymbol{d} : \{d_A, d_B\} = \{2, 2\}$ in (a) and $\{1, 1\}$ in (b).

maximum height, and $\boldsymbol{d}$ *is the set of multi-homing factor for all nodes* $n \in \boldsymbol{\mathcal{N}}$.

Figure 4.2a and 4.2b show examples of how the parameters h and q are set for different traffic and topologies.

4.4 Fair Price Achieving the Shapley Value: The Symmetric Case

The Shapley value is known as the fair way to distribute contributions among a group of players. However, adopting Shapley value in the current Internet is challenging since it requires the ISPs to first form a coalition, and besides the Shapley settlement requires centralized computation. Nevertheless, before we contemplate the likelihood of such a development, it would be helpful to find out the possibility of achieving the same revenue distribution as *Shapley* through the current Bilateral scheme, under suitably chosen prices.

In this section, we calculate the traffic-based price for each tier such that *Bilateral* produces the same revenue distribution as *Shapley*. We define the set of corresponding prices $\boldsymbol{P}^*$ as the *fair prices*:

Definition 4.4.1 (Fair price). *The fair price* $\boldsymbol{P}^* = \{P_t^*, t \in [1, \ldots, T-1]\}$ *is the set of tier-based prices that make the expected revenue distribution in* Bilateral *equal to* Shapley *for all tiers* T, *i.e.,*

$$\begin{aligned} &P_a \sum_{n \in \mathcal{N}_t} \sum_{(a,b) \in \mathbf{R}} \mathbb{P}(a,b) \varsigma_n(a,b) \\ &= \sum_{n \in \mathcal{N}_t} \sum_{(a,b) \in \mathbf{R}} \mathbb{P}(a,b) \gamma_{n,t}(\boldsymbol{P}^*) + P_a \times \mathbf{1}_{\{t=T\}}, \forall t \in \{1, \ldots, T\}, \end{aligned}$$

where $\mathbf{R}$ *is the set of all possible source-destination pairs,* $\mathbb{P}(a,b)$ *is the probability that traffic (a,b) happens, and* $\mathbf{1}_{\{.\}}$ *is the indicator function.*

In the rest of the paper, we will consider how the fair prices can be calculated in both symmetric and asymmetric topologies, as well as some nice properties of the fair prices (using symmetric topology as an example). After that, we will consider how the fair prices can be approximated in a distributed fashion based on only local information.

We first consider a symmetric topology, $\mathcal{G}^s = (T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}^s, d)$, where $\boldsymbol{\rho}^s = \{\rho_t, t \in T\}$. In this topology, each node in the same tier t shares the same peering parameter ρ_t and all nodes of all tiers share the same multi-homing parameter d. For traffic parameters, we assume all ISPs have the same traffic parameter α, the total traffic intensity in the network is normalized to 1.

Theorem 4.4.2. *The fair price* $\boldsymbol{P}^*(\mathcal{G}^s)$ *in a symmetric hierarchical topology* $\mathcal{G}^s = (T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}^s, d)$ *is:*

$$P_t^*(\mathcal{G}^s) = \frac{P_a}{2(1-\alpha\rho_{t+1})} \left\{ \Gamma_{\mathcal{G}^s}^{tier1} + \Gamma_{\mathcal{G}^s}^{no\text{-}peer} + \Gamma_{\mathcal{G}^s}^{peer} \right\}$$

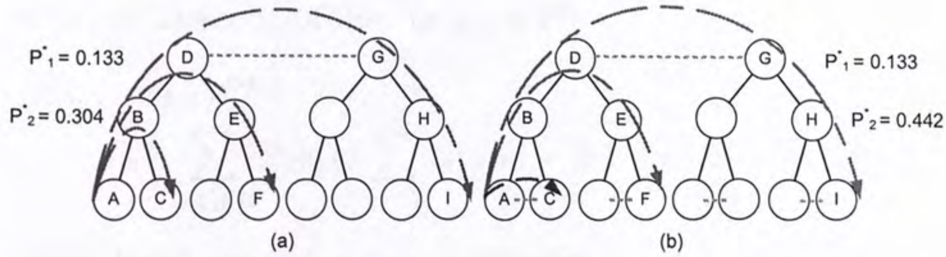

Figure 4.3: Numerical examples of calculating fair price. We consider a three-tier model, the end-users are omitted in the figure. We let $P_a = 1$, $\alpha = 0.5$. Without loss of generality we draw all possible traffic (in terms of height travelled) initialized by node A by dash arrows. (a) shows the fair prices when $\rho_3 = 0$ and (b) show the fair prices when $\rho_3 = 1$.

$$
\begin{aligned}
&\forall t \in \{1, \ldots, T-1\}, where \\
&\qquad \Gamma_{\mathcal{G}^s}^{tier1} = 2t(1-\alpha)^t \psi_t(T, 1, d), \\
&\qquad \Gamma_{\mathcal{G}^s}^{no\text{-}peer} = \sum_{i=1}^{t} (2i-1)\alpha(1-\alpha)^{i-1}(1-\rho_{t+2-i}) \times \\
&\qquad\qquad \psi_t(T-t+i, 0, d), \\
&\qquad \Gamma_{\mathcal{G}^s}^{peer} = \sum_{i=1}^{t-1} 2i\alpha(1-\alpha)^i \rho_{t+1-i}\, \psi_t(T-t+i, 1, d).
\end{aligned}
$$

This theorem is proofed by dynamic programming, which is omitted in this paper. Here $\Gamma_{\mathcal{G}^s}^{\text{tier1}}$ is the Shapley value corresponding to the traffic passing through the peered links in mesh at tier 1 times the probability of these traffic happens. This term is unique since we have assumed that the tier 1 nodes form a full mesh network through peering links with probability 1, thus is different from all other tiers. $\Gamma_{\mathcal{G}^s}^{\text{no-peer}}$ is related to the Shapley value of all traffic going higher than tier t but do not use peering links at the highest height. $\Gamma_{\mathcal{G}^s}^{\text{peer}}$ is related to the Shapley value of traffic that passes through peering links at the traffic's highest height.

Through rearranging the equations in Theorem (4.4.2), we obtain the following relationship that sheds more light on the economic meaning of the fair price $\boldsymbol{P}^*$:

$$2(1-\alpha\rho_{t+1})P_t^* = \\ P_a\Big\{\sum_{(a,b)\in\mathbf{R}} \mathbb{P}(a,b)\sum_{n\in N_t} \varsigma_n(a,b) + 2(1-\alpha)(1-\alpha\rho_t)P_{t-1}^*\Big\},$$

where the term $2(1-\alpha\rho_{t+1})$ denotes the total traffic intensity(both upstream and downstream traffic) between all ISPs at tier t and tier $t+1$. The first term in the braces is the Shapley value of all ISPs at tier t considering all traffic in the network. The second term in the braces is the Shapley value of the higher tiers (from tier 1 to $t-1$). The term $2(1-\alpha)(1-\alpha\rho_t)$ represents the total traffic between tier t ISPs and their provider ISPs, and the fair price $P_{\mathcal{N}_{t-1}}^*$ is the price that makes this bilateral settlement equals to Shapley value at tier $t-1$. As a result, the fair price at tier t is to find the per-traffic price that accounts the contribution of tier t, plus the price that contains the contribution of the upper tiers so that the upper tiers can achieve *Shapley* by charging tier t ISPs $P_{\mathcal{N}_{t-1}}^*$. From the second term of the braces, we can see that long distance routes are more expensive to transport(i.e. need to be charged at higher prices). It is because when the path is longer, more revenue is needed to carry the portion that belongs to the rest of the contributors in the route.

By this formulation, the fair price at tier t is to account the contribution of tier t ISPs (first term on the right), plus the price that contains the contribution of the upper tiers (second term on the right) so that the upper tiers can achieve *Shapley* by charging tier t ISPs P_{t-1}^*. We then obtain the per-traffic based fair price by dividing the contributions by the total traffic intensity between all ISPs at tier t and tier $t+1$ (the term on the left).

Figure 4.3 shows a numerical example of calculating the fair prices. When peering links are established at tier 3 in Figure

4.3b, the fair price of tier 2 increases. When $\rho_3 = 1$, shorter routes like A to C will not pass through tier 2 ISP B. Since the fair price is the average per-traffic price of handling a traffic, losing the relatively cheap short distance traffic raises the average per-traffic price. Note that the actual revenue of tier 2 has decreased when $\rho_3 = 1$ as they handle less traffic.

Since using fair price in bilateral settlement helps obtaining Shapley value, which is the fair allocation of revenues based on different nodes' contribution, we could use this fair price as one of the factors to determine the reasonable per-traffic price in the negotiation of bilateral agreements.

4.5 Properties of the Fair Prices in the Symmetric Case

Under fair prices, *Bilateral* achieves the same revenue distribution as *Shapley*. Next we show that fair prices exhibits nice properties in the system parameters.

4.5.1 Sensitivity to traffic pattern α

Proposition 4.5.1. *The fair price $\boldsymbol{P}^*$ is decreasing in $\alpha \in (0, 1]$ for any $t \in [1, \ldots, T-1]$.*

The proposition is proofed by mathematical induction on the derivative of the fair price on α, which is not shown in this paper. Figure 4.4 shows the relationship between the fair price $\boldsymbol{P}^*$ and α by calculating the fair prices across α. When α increases, the traffic is more likely to go local, and thus is more likely to be handled by tiers that are closer to the leaf tiers. As a result, ISPs of tiers higher on the hierarchy have less chance to contribute to the connectivity of the traffic, thus it makes sense for them to charge less. On the other hand, when α decreases, a traffic tends to travel further away. In the extreme case where

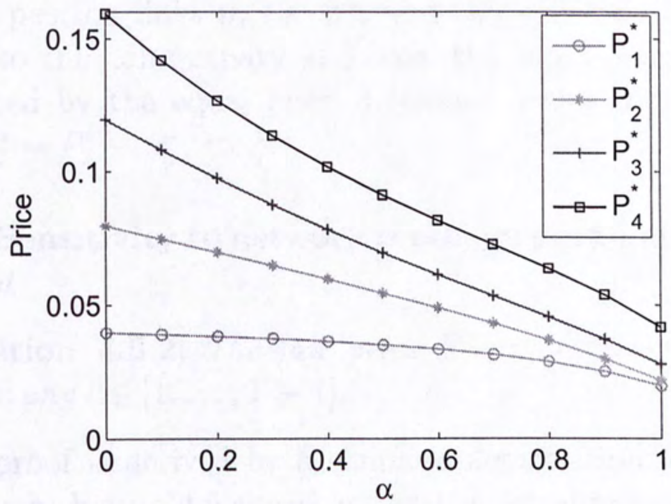

Figure 4.4: Fair price P_t^* verses α for different tier t: the fair price is decreasing in α.

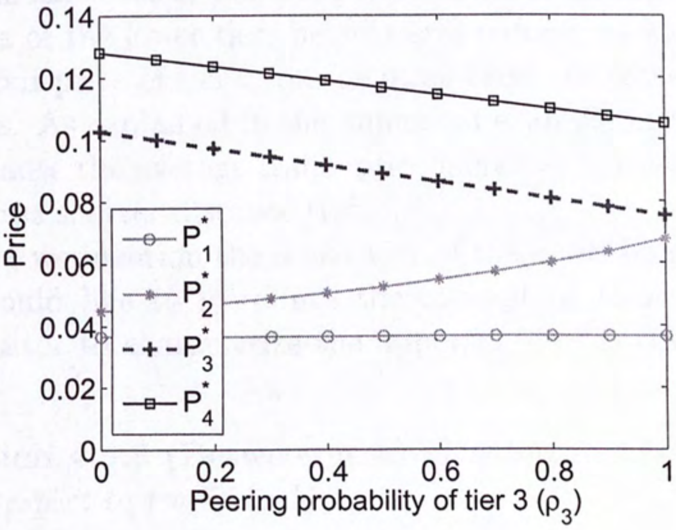

Figure 4.5: P_t^* verses peering probability of tier 3 (ρ_3) for different tier t: the fair price of tier t is decreasing in ρ_t

α approaches 0, all traffic travel to the farthest distance (i.e., through peering links in tier 1), and thus all tiers contribute equally to the connectivity and have the same revenue. This is reflected by the equal price difference across different tiers ($P_1^* - P_2^* = P_2^* - P_3^*$, etc.).

4.5.2 Sensitivity to network topology parameters ρ and d

Proposition 4.5.2. *The fair price* $\boldsymbol{P}^*$ *is decreasing in* $\rho_t \in (0,1)$ *for any* $t \in [1, \ldots, T-1]$.

The proof is derived by a simple differentiation that is not shown here. Figure 4.5 shows an mathematical example of how the fair prices of different tiers change as a function of the peering probability of tier 3 ISPs ρ_3. When ρ_3 increases, more traffic passes through the peering link in tier 3 without using services from their provider ISPs. Paying less to the upstream ISPs reduces the fair price of tier 3 ISPs. With the same argument, the fair price of the lower tiers below tier 3 reduces as well.

The fair price of tier 2, on the other hand, increases when ρ_3 increases. As explained in the numerical example in Figure 4.3, it is because the average traffic price increases when tier 2 ISPs handle less shorter distance traffic.

Before we mention the sensitivity of the multi-homing factor d, we could like to introduce the concept of *betweenness*, as an indicator to characterize the importance of an ISP node as follows:

Definition 4.5.3 (Betweenness). *The betweenness of a node* n *with respect to traffic* (a, b) *is:*

$$\mathcal{B}_n(a,b) = \frac{\sigma(a,n,b)}{\sigma(a,0,b)},$$

where $\sigma(a, n, b)$ is the number of routes from a to b that pass through node n. If $n = 0$, then $\sigma(a, 0, b)$ denotes the total number of routes from a to b. It is clear that $\mathcal{B}_a(a, b)$ and $\mathcal{B}_b(a, b)$ are always equal to 1.

Observation 4.5.4. *The Shapley value of an ISP n increases in multi-homing factor d if its betweenness $\mathcal{B}_n = 1$, and decreases in d if its betweenness $\mathcal{B}_n < 1$. Furthermore, $\boldsymbol{P}^* \to 0$ when $d \to \infty$.*

Figure 4.6 shows the differences in Shapley value between a vital player (betweenness $\mathcal{B} = 1$) and a non-vital player ($\mathcal{B} < 1$) for varies d. When d increases, the Shapley value of non-vital ISPs decreases significantly since there are more alternative ways for providing connectivity, while the Shapley value of vital ISPs increases significantly since they remain the necessary components for any choice of connectivity.

Figure 4.7 shows that $\boldsymbol{P}^*$ drops dramatically when d increases and flattens quickly for $d \geq 5$, as this connectivity-focused characteristic function gives little value to non-vital ISPs. For $d \geq 1$, the only vital players for each traffic are the initiating ISP and the terminating ISP. In this scenario, initiating leaf ISP and the terminating leaf ISP at tier T collect most of the revenue from handling the traffic, leaving the rest of the non-vital upper tier ISPs little revenue to share. This explains why we have $\boldsymbol{P}^* \to 0$ when $d \to \infty$. In this extreme case, there are infinite number of paths that can be used to transmit the same traffic, thus the contribution of a non-vital ISP on any particular path to connectivity is virtually zero.

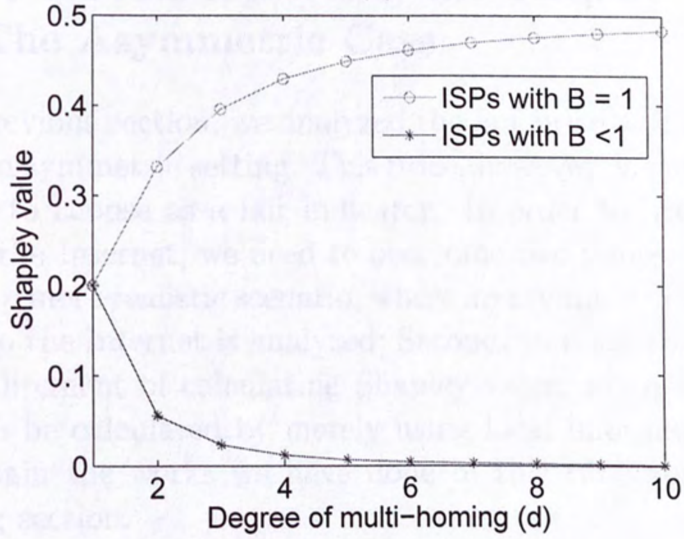

Figure 4.6: Shapley value ς for different multi-homing factor d for vital and non-vital ISPs: When d increases, vital nodes raises its contribution and non-vital nodes drops its contribution.

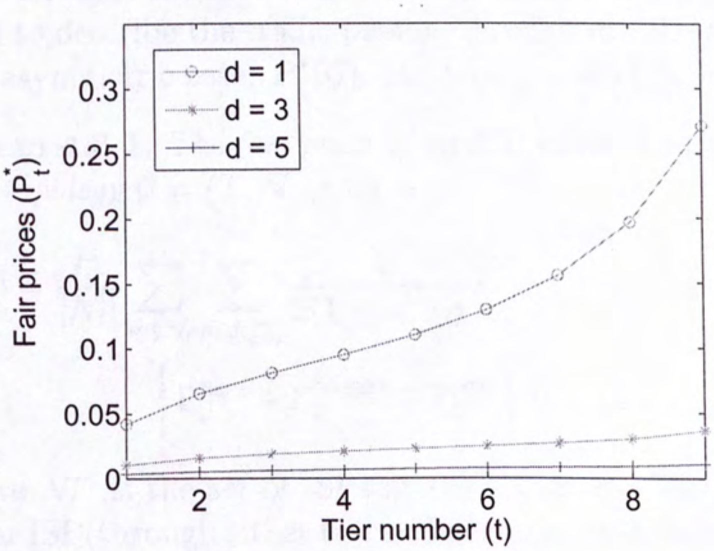

Figure 4.7: $\boldsymbol{P}^*$ for different multi-homing factor d: The fair price drops drastically when d increases.

4.6 Fair Price Achieving the Shapley Value: The Asymmetric Case

In the previous section, we analyzed the fair price and its implications in symmetric setting. This price, however, is not feasible for ISPs to choose as a fair indicator. In order to use the fair price in the Internet, we need to overcome two issues: First, to consider a more realistic scenario, where an asymmetric topology similar to the Internet is analyzed; Second, to relax the centralized requirement of calculating Shapley value, so that the fair price can be calculated by merely using local information. We will explain the works we have done in this two areas in the following section.

Here we relax the symmetric network assumption in Sections 4.5 and 4.4, and show how the fair price is determined in the asymmetric case. Here we consider a general topology model $\mathcal{G} = (T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}, \boldsymbol{d})$, where $\boldsymbol{\rho} \in \{\rho_n, \forall n \in \boldsymbol{\mathcal{N}}\}$ and $\boldsymbol{d} \in \{d_n, \forall n \in \boldsymbol{\mathcal{N}}\}$. In terms of traffic model, each ISP node n uses parameters (α_n, I_n) to describe the traffic passing through it. The *fair price* in this asymmetric case, $\boldsymbol{P}^*(\mathcal{G})$, can be calculated as follows:

Theorem 4.6.1. *The fair price of an ISP at tier t in an asymmetric topology $\mathcal{G} = (T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}, \boldsymbol{d})$ is:*

$$P_t^*(\mathcal{G}) = \frac{P_a}{|\mathcal{N}_t|} \sum_{m \in \mathcal{N}_t} \sum_{n \in \mathcal{N}_{t+1}^m} \frac{I_n}{2(1 - \alpha_n \rho_n)} \times \left\{ \Gamma_{\mathcal{G}}^{tier1} + \Gamma_{\mathcal{G}}^{no\text{-}peer} + \Gamma_{\mathcal{G}}^{peer} \right\} \forall t \in \{1, \ldots, T-1\},$$

where $\mathcal{N}_t^m$ is the set of ISPs at tier t that has ISP m as its provider ISP(through either transit links or multi-homing links). The three pricing terms in Theorem (4.6.1) can be similarly derived as in Theorem (4.4.2), and is not shown in this paper.

4.7 Distributed and Local Approximation of the Fair Price

In order to calculate the fair prices based on either Theorem (4.4.2) (symmetric case) or Theorem (4.6.1) (asymmetric case), we require global information such as the full network topology, which is impractical to obtain in the Internet. Next we illustrate a distributed approximation scheme that allows each node to calculate the fair price with limited local information.

For a node n at tier t, it calculates an approximated fair price $\boldsymbol{P'^*}$ based on the following inputs: 1) the price charged by its provider ISPs at tier $t-1$ (P^*_{t-1}); 2) the incoming traffic pattern $(\alpha_m, I_m), \forall m \in \mathcal{N}_t$; 3) the betweenness $\mathcal{B}_m$ of all nodes $m \in \mathcal{N}_t$, and 4) a heuristic approximation of peering parameters $\rho'_t, t \in \{2, \ldots, t-1\}$.

The fair price can be approximated by

$$P'^*_t(\mathcal{G}) = \frac{P_a}{|\mathcal{N}_t|} \sum_{m \in \mathcal{N}_t} \sum_{n \in \mathcal{N}^m_{t+1}} \frac{I_n}{2(1-\alpha_n \rho_n)} \left\{ \Gamma'^{\text{tier1}}_{\mathcal{G}} + \Gamma'^{\text{guess}}_{\mathcal{G}} + \Gamma'^{\text{tier}t}_{\mathcal{G}} + \Gamma^{\text{Bilateral}}_{\mathcal{G}} \right\}, \quad (4.1)$$

where $\Gamma'^{\text{tier1}}_{\mathcal{G}}$ and $\Gamma'^{\text{tier}t}_{\mathcal{G}}$ are the approximated Shapley value of traffic to tier 1 ISPs and tier t ISPs respectively. $\Gamma'^{\text{guess}}_{\mathcal{G}}$ is the approximated Shapley value of traffic that goes to heights between tier 2 and tier $t-1$, which requires ρ'_t for approximation. $\Gamma^{\text{Bilateral}}_{\mathcal{G}}$ is the precise price that tier t ISPs pay their provider ISPs for all upstream traffic. Description of terms of Equation (4.1) is shown in Appendix A.2.3

This approximation is based on two observations. First, we approximate the original Shapley value function $\psi_n(T, q, \boldsymbol{d})$ by $\psi'_n(T, \mathcal{B}_n)$, which requires only local information to compute. We can think of the betweenness $\mathcal{B}$ as the number of competitors of node n. The more multi-homing links from the subscriber

ISPs of node n, the smaller value of betweenness of node n. The second observation is that a general idea of how each tier peers is feasible to obtain, and sufficient to calculate a close fair price. As shown in Figure 4.5, the fair prices do not significantly change for two close ρ_t, therefore we can use a heuristic approximation to replace this piece of information.

The limited information from the above tier and all peers within the same tier enables us to precisely calculate $\Gamma_{\mathcal{G}}^{\text{tier1}}$ (Shapley value obtained by tier-1 nodes), $\Gamma_{\mathcal{G}}^{\text{tier}t}$(Shapley value obtained by tier t nodes), and $\Gamma_{\mathcal{G}}^{\text{Bilateral}}$(price charged from the upper tiers in *Bilateral*). The term $\Gamma_{\mathcal{G}}^{\text{guess}}$(Shapley value of tier 2 to tier $t-1$ ISPs) is calculated based on a heuristic approximation of the peering pattern of other tiers.

To illustrate the performance of the approximation, we simulate a four-tier topology where the real peering parameter and heuristic approximation of the peering pattern are both independently and randomly generated. We use the following approximation for the Shapley value calculation:

$$\psi_n'(T,\mathcal{B}_n) = \frac{1}{2}\Big[\psi_n(T,1,\boldsymbol{d'}_n) + \psi_n(T,0,\boldsymbol{d'}_n)\Big],$$

where $\boldsymbol{d'}_n = \{3,\dots,d_n,\dots,3\}$, i.e., assuming all tiers have the same multi-homing factor equal to 3, except the value at the $t+1$ tier which is approximated by d_n.

Figure 4.8 shows that the approximation is very accurate in this random case, where the difference is no larger than 8% (Figure 4.9). If the peering parameters are estimated according to some network statistics instead of just random guessing, then the performance of the approximation is likely to be further improved.

□ **End of chapter.**

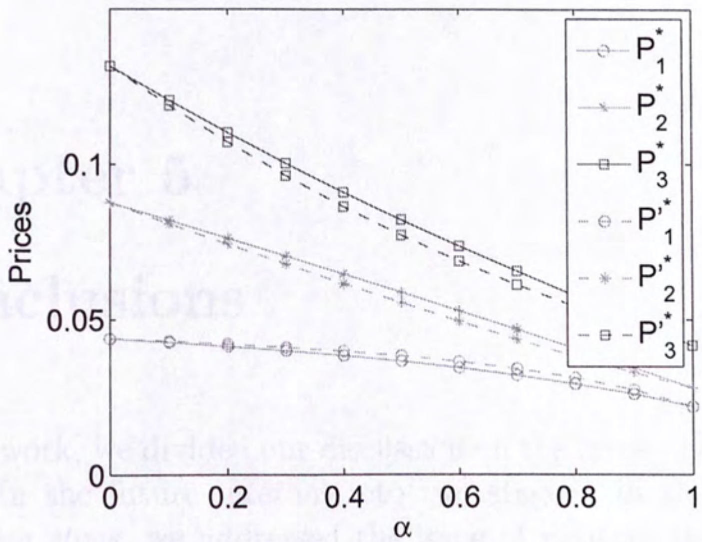

Figure 4.8: Simulated performance of the distributed approximation scheme for the fair price. Here we consider a four-tier topology with randomly chosen true and approximated peering parameters.

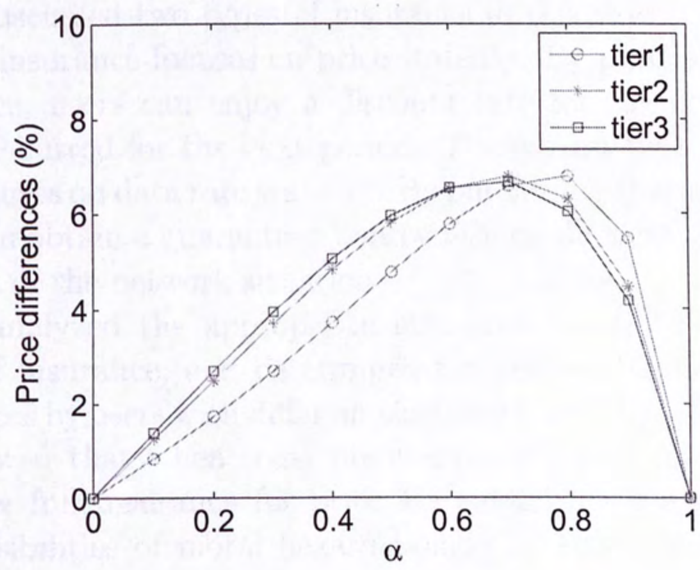

Figure 4.9: The difference between real and approximated fair prices of different tiers in percentage for Figure 4.8

Chapter 5

Conclusions

In this work, we divided our discussion on the economic revenue model in the future Internet into two stages. In the *revenue collecting stage*, we addressed the issue of network inefficiency in the current Internet. We analyzed how the concept of insurance can be introduced to Internet pricing, such that the novel pricing mechanisms, for instance congestion pricing, are feasible to implement in both users' and ISP's perspective.

We discussed two types of insurance in this work. The first type of insurance focuses on price stability. By purchasing this insurance, users can enjoy a discount rate for the congestion charge incurred for the next period. The second type of insurance focuses on data rate stability. By purchasing this insurance, users can obtain a guaranteed bandwidth for the next period regardless of the network situation.

We analyzed the appropriate insurance functions for both types of insurance, and determined the optimal choices of the insurances by users with different elasticities and risk adversities. We showed that when some premium is required for users to purchase full insurance for price stability, ISPs can minimize the possibilities of moral hazard bought to the system. Also, by correctly setting the per-unit bandwidth price for data rate stability, we are able to provide guaranteed bandwidth service to users with important applications, and on the other hand help

ISPs to break even when there is no congestion in the network.

For future work in this section, we can discuss the practical issues of implementing insurance in the Internet. We can start the analysis by checking the existing pricing contracts in the networking/telecommunication business to gain more insights on how the concept of insurance can be introduced. In another direction, we can verify this mechanism for distributed, large scale environment such as the Internet, and to determine the algorithm to perform the insurance mechanism.

The second stage we mentioned in this work is the *revenue distributing stage*. We showed that the current bilateral method to distribute revenue does not distribute revenue fairly according to network players' contribution. We try to answer the following question: is it possible to set the proper bilateral prices so that a fair revenue distribution among ISPs can be achieved using the revenue settlement mechanism used in the current Internet? Through both careful analysis and extensive simulations, we have obtained positive preliminary answers to this question. We show that by setting the bilateral charging based on global topology and traffic information for each ISP tier, the current bilateral settlement distribution leads to Shapley value, which is widely regarded as a fair cooperative revenue distribution method. We further show that such fair prices are reasonable to be adopted in the Internet, as it shows sensible properties to parameter changes. In terms of feasibility to be implemented, we show that the fair price can be approximated using local information.

There are several possible directions of extending the current work in this section. One direction is to consider different characteristic functions and analyze the properties of Shapley settlement in that case. The other direction is to study the incentive for ISPs to implement the Shapley settlement in realistic network scenarios.

□ **End of chapter.**

Appendix A

Mathematical Proofs

A.1 Mathematical Proof for Chapter 3

A.1.1 Proof of Theorem 3.3.2

We proof this theorem by applying the definition of actuarially fair in Definition 3.3.1. The indemnification is the portion of congestion charge that is being insured and is paid by the insurer. In this case, the indemnification is $(1-r) \times \frac{\sum_{t\in T} x_n(t)p(t)}{\sum_{t\in T} X(t)p(t)} cY$.

$$\rho\Big(m - (1-r)\frac{\sum_{t\in T} x_n(t)p(t)}{\sum_{t\in T} X(t)p(t)}cY\Big) + (1-\rho)(m)$$

$$r = 1 - \frac{m}{\rho\frac{\sum_{t\in T} x_n(t)p(t)}{\sum_{t\in T} X(t)p(t)}cY}$$

$$r = 1 - \frac{m}{\rho E}.$$

A.1.2 Proof of Proposition 3.3.5

We let the maximization problem of a risk-averse user in Equation (3.1) be g and denote the expected utility function $V^{RA}(\cdot)$. As $V^{RA}(\cdot)$ is concave, the maximization problem can be solved by setting the first derivative of g to zero:

$$\begin{aligned} g &= \max_m \rho V^{RA}(w - m - r(m)E) + (1-\rho)V^{RA}(w-m) \\ &= \max_m \rho V^{RA}(w - m - (1 - \frac{m}{E})E) + (1-\rho)V^{RA}(w-m) \\ \frac{\partial g}{\partial m} &= \rho(-1 + \frac{1}{\rho})V^{RA'}(w - m - (1 - \frac{m}{E})E) \\ &\quad - (1-\rho)V^{RA}(w-m) = 0, \end{aligned}$$

the second line of the deviation is obtained by substituting the coinsurance function $r(m)$ calculated in the previous subsection. As $V^{RA}(\cdot)$ is concave, strictly increasing and unique, when $V'_{RA}(x) = V'_{RA}(y)$, essentially $x = y$. We then obtain:

$$\begin{aligned} w - m - (1 - \frac{m}{E})E &= w - m \\ m^* &= E \end{aligned}$$

A.1.3 Proof of Proposition 3.3.6

We denote the expected utility function of a risk-seeking user be $V^{RS}(\cdot)$. As we can observe that the maximization problem in Equation (3.1) is to find the utility value in between $V^{RS}\big(w - m - r(m)E\big)$ and $V^{RS}(w-m)$. Intuitively, as $V_{RS}(\cdot)$ is convex, the maximum value of Equation (3.1) happens at the positive boundary of the utility function. Therefore the best response of risk-seeking users is to maximize $V^{RS}(w-m)$, that is to set $m^* = 0$.

A.1.4 Proof of Proposition 3.3.7

Since risk neutral users do not have special preference to risk and actuarially fair rate gives the same mean payment for users with different insurance charge, it is not necessary for risk neutral users to purchase any insurance, or in other words, they have the same utility surplus by any insurance charge m.

We denote the expected utility function of a risk-neutral user be $V^{RN}(\cdot)$. As $V^{RN}(\cdot)$ is a linear function, we generalize the function as $V^{RN}(x) = \delta x$ to illustrate the above concept:

$$\begin{aligned}
&\max_m \ \rho\delta\Big(w - m - r(m)E\Big) + (1-\rho)\delta(w-m) \\
\Rightarrow\ &\max_m \ \delta w - \delta m - \rho\delta r(m)E \\
\Rightarrow\ &\max_m \ \delta w - \delta m - \rho\delta(1 - \frac{m}{E})E \\
\Rightarrow\ &\max_m \ \delta w - \rho\delta E
\end{aligned}$$

As the maximization problem is independent of the insurance payment m, the optimal insurance charge m^* can be any value in $(0, E)$.

A.1.5 Proof of Proposition 3.4.1

The proof follows the fact that the demand function $D(\cdot)$ is a monotonic decreasing and convex function. Therefore, the maximum difference happens when $r(m)$ is minimal. That is when $r(m) = 0$, which means $m^* = E$.

A.1.6 Proof of Proposition 3.4.3

We let the maximization problem in Equation (3.1) be g'. By the concavity of the expected utility function of risk-averse users, we can solve g' by setting the first derivative of g' to zero.

$$
\begin{aligned}
g' &= \max_m \rho V^{RA}(w - m - r^p(m)E) + (1-\rho)V'^{RA}(w-m) \\
\frac{\partial g'}{\partial m} &= \rho(-1 + \frac{\alpha}{\rho})V'^{RA}(w - m - r^p(m)E) \\
&\quad - (1-\rho)V'^{RA}(w-m) = 0,
\end{aligned}
$$

we let $V^{RA}(x) = \log(x)$ as an generic example of a concave function to continue the calculation. The optimal insurance price m^p under the premium coinsurance function r^p and a specific utility function is:

$$
m^{p*} = \frac{\rho\big[w(\alpha - 1) + E(1-\rho)\big]}{\alpha - \rho}.
$$

the derivative of m^{p*} with respect to α is:

$$
\begin{aligned}
\frac{\partial m^{p*}}{\partial \alpha} &= \frac{\rho w}{\alpha - \rho} - \frac{2\rho w \alpha}{(\alpha-\rho)^2} \\
&= \frac{\rho w}{\alpha \rho}\left[1 - \frac{2\alpha}{\alpha - \rho}\right] \\
&< 0,
\end{aligned}
$$

given that $0 \leq \alpha \leq 1$, $0 \leq \rho \leq 1$, and $w > 0$. Note that the function of m^{p*} is not continuous.

A.1.7 Proof of Proposition 3.4.5

It is intuitive that risk-seeking users do not purchase any insurance under premium discount rate, given that risk-seeking users do not purchase any insurance even under actuarially fair rate.

For risk neutral users, as $V^{RN}(\cdot)$ is linear, we generalize the function as $V^{RN}(x) = \delta x$ to illustrate their optimal insurance payment:

$$\begin{aligned}
&\max_m \ \rho\delta(w - m - \sum_{t \in T} r(m)E) + (1-\rho)\delta(w-m) \\
\Rightarrow &\max_m \ \delta w - \delta m - \rho\delta(1 - \frac{m\alpha}{E})E \\
\Rightarrow &\max_m \ \delta w - \rho\delta E - \delta m(1-\alpha)
\end{aligned}$$

Since $\alpha \in (0,1)$ and $m \geq 0$, the optimal response of insurance charge m'^* for risk-neutral users are $m'^* = 0$.

A.2 Mathematical Proof for Chapter 4

A.2.1 Proof of Theorem 4.4.2

The proof is based on mathematical induction. First, we calculate the fair price of tier 1 as

$$P_1^*(\mathcal{G}^s) = \frac{P_a\Big\{2(1-\alpha)\psi_1(T,1,d) + (1-\rho)\alpha\psi_1(T,0,d)\Big\}}{2(1-\alpha\rho)}. \tag{A.1}$$

From tier 2 onward, the revenue of the bilateral settlement method requires the pricing information of the upper tier P_{t-1}^*. Therefore the fair price for tier t is:

$$\begin{aligned}
P_t^*(\mathcal{G}^s) = \frac{P_a}{2(1-\alpha\rho_{t+1})}\Bigg\{&(1-\alpha)^t \sum_{\phi \in \mathcal{N}_t} \psi_\phi(T,1,d) + \\
&\sum_{i=1}^{t} \alpha(1-\alpha)^{i-1}(1-\rho_{t+2-i}) \sum_{\phi \in \mathcal{N}_t} \psi_\phi(T-t+i,0,d) + \\
&\sum_{i=1}^{t-1} \alpha(1-\alpha)^i \rho_{t+1-i} \sum_{\phi \in \mathcal{N}_t} \psi_\phi(T-t+i,1,d) + \\
&2(1-\alpha)(1-\alpha\rho_{t+1})P_{t-1}^*\Bigg\}.
\end{aligned} \tag{A.2}$$

The value inside the braces (except the term with P_{t-1}^*) is the expected Shapley value of ISP nodes through handling the traffic injected into tier t, which includes all traffic goes through any nodes in tier t. We then solve this recursive equation, and obtain the close-form solution of the fair price.

$$P_t^*(\mathcal{G}^s) = \frac{P_a}{2(1-\alpha\rho_{t+1})}\Big\{(1-\alpha)^t \sum_{i=1}^{t}\sum_{\phi\in\mathcal{N}_i}\psi_\phi(T,1,d)+$$
$$\sum_{i=1}^{t}\alpha(1-\alpha)^{i-1}(1-\rho_{t+2-i})\sum_{j=1}^{i}\sum_{\phi\in\mathcal{N}_{t-j+1}}\psi_\phi(T-t+i,0,d)+$$
$$\sum_{i=1}^{t-1}\alpha(1-\alpha)^{i}\rho_{t+1-i}\sum_{j=1}^{i}\sum_{\phi\in\mathcal{N}_{t-j+1}}\psi_\phi(T-t+i,1,d)\Big\}. \quad (A.3)$$

Since we assume that a multi-homing link can provide connectivity for all traffic from its subscribers, and each ISPs has d multi-homing links. Then the betweenness $\mathcal{B}$ of each ISP node n at all tiers other than the leaf tier is all ISPs at each tier is $\frac{1}{d}$. For a leaf tier ISP, the betweenness equals to 1. Therefore we are able to simplify Equation (A.3) by grouping the same terms, which leads to the final representation in the theorem.

A.2.2 Proof of Theorem (4.6.1)

The terms in the theorem are as follows:

$$\begin{aligned}
\Gamma_{\mathcal{G}}^{\text{tier1}} =& (1-\alpha_n)^t \sum_{i=1}^{t} \sum_{\phi \in \mathcal{N}_i} \psi_\phi(T, 1, \boldsymbol{d}), \\
\Gamma_{\mathcal{G}}^{\text{no-peer}} =& \sum_{i=1}^{t} \alpha_n (1-\alpha_n)^{i-1} \sum_{j=1}^{i} \sum_{\phi \in \mathcal{N}_{t-j+1}} (1-\rho_\phi) \times \\
& \psi_\phi(T-t+i, 0, \boldsymbol{d}), \\
\Gamma_{\mathcal{G}}^{\text{peer}} =& \sum_{i=1}^{t-1} \alpha_n (1-\alpha_n)^{i} \sum_{j=1}^{i} \sum_{\phi \in \mathcal{N}_{t-j+1}} \rho_\phi \psi_\phi(T-t+i, 1, \boldsymbol{d}).
\end{aligned}$$

The key idea here is to find the average fair prices of all possible traffic sources transversing through all possible subsets of the topology. For each ISP m at tier $t \neq T$, it has a set of subscriber ISPs, either established by transit links or multi-homing links. We denote this set of subscribers as $\mathcal{N}_{t+1}^m$. For each subscriber ISP $n \in \mathcal{N}_{t+1}^m$, it has a specific set of traffic parameters (α_n, I_n) experienced by ISP node m. Therefore the fair price from node m's perspective is:

$$P_t^*(\mathcal{G}_m^s) = \sum_{n \in \mathcal{N}_{t+1}^m} I_n P_t^*(\mathcal{G}_m^s(n)), \tag{A.4}$$

where $\mathcal{G}_m^s = (T, \boldsymbol{\mathcal{N}}, \boldsymbol{\rho}^s, \boldsymbol{d})$ is the symmetric topology that node m thinks about the whole topology based on its own local information (although the actual topology is asymmetric). We calculate $P_t^*(\mathcal{G}_m^s(n))$, the fair price of the traffic from node n, by Equation (A.3). As in this asymmetric setting, each node has its own multi-homing factor, thus the Shapley value of individual nodes are typically different.

The final fair price of tier t in this asymmetric case would be the value that averaged over all nodes in tier t, i.e.,

$$P_t^*(\mathcal{G}) = \frac{1}{|\mathcal{N}_t|} \sum_{m \in \mathcal{N}_t} P_t^*(\mathcal{G}_m^s). \tag{A.5}$$

The final expression in Theorem (4.6.1) can be obtained by combining (A.4) and (A.5).

A.2.3 Terms Description of Equation (4.1)

The terms of the equation are as follows:

$$\begin{aligned}
\Gamma_{\mathcal{G}}^{\prime\text{tier1}} =& (1-\alpha_n)^t \sum_{\phi \in \mathcal{N}_t} \psi'_\phi(T, \mathcal{B}_\phi), \\
\Gamma_{\mathcal{G}}^{\prime\text{guess}} =& \sum_{i=1}^{t-1} \alpha_n (1-\alpha_n)^{i-1} \sum_{\phi \in \mathcal{N}_t} [1 + (1-\alpha_n)\rho'_\phi - \rho'_\phi] \times \\
& \psi'_\phi(T - t + i, \mathcal{B}_\phi), \\
\Gamma_{\mathcal{G}}^{\prime\text{tier}t} =& \alpha_n (1-\alpha_n)^{t-1} \sum_{\phi \in \mathcal{N}_t} (1 - \rho^\phi) \psi'_\phi(T, \mathcal{B}_\phi), \\
\Gamma_{\mathcal{G}}^{\text{Bilateral}} =& 2(1-\alpha_n)(1-\alpha_n \rho_t) P_{t-1}^*.
\end{aligned}$$

The above equations are derived from Equation (A.2), after taking into consideration of the approximation of Shapley value (from $\psi_n(T, q, \boldsymbol{d})$ to $(\psi'_n(T, \mathcal{B}_n))$ and the approximation of the peering parameters of other tiers.

□ **End of chapter.**

Bibliography

[1] L. Anania and R. Solomon. Flat–the minimalist price. *Internet Economics*, pages 91–118, 1997.

[2] E. Anderson, F. Kelly, and R. Steinberg. A Contract and Balancing Mechanism for Sharing Capacity in a Communication Network. *MANAGEMENT SCIENCE*, 52(1):39, 2006.

[3] J. Bailey. The Economics of Internet Interconnection Agreements. *Internet Economics*, pages 155–168, 1995.

[4] D. Bertsekas and R. Gallager. *Data networks.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1987.

[5] B. Briscoe. Re-ECN: Adding Accountability for Causing Congestion to TCP/IP. *draft-briscoe-tsvwg-re-ecn-tcp-03 (work in progress), October*, 2006.

[6] B. Briscoe. Flow rate fairness: dismantling a religion. *ACM SIGCOMM Computer Communication Review*, 37(2):63–74, 2007.

[7] B. Briscoe, A. Jacquet, C. Di Cairano-Gilfedder, A. Salvatori, A. Soppera, and M. Koyabe. Policing congestion response in an internetwork using re-feedback. *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 277–288, 2005.

[8] D. Chiu and R. Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN Systems*, 17(1):1–14, 1989.

[9] D. Clark. Internet cost allocation and pricing. *Internet Economics*, pages 95–112, 1997.

[10] D. Cutler and R. Zeckhauser. The Anatomy of Health Insurance. *NBER WORKING PAPER SERIES*, 1999.

[11] R. Edell and P. Varaiya. Providing Internet access: what we learn from INDEX. *Network, IEEE*, 13(5):18–25, 1999.

[12] M. Falkner, M. Devetsikiotis, and I. Lambadaris. An overview of pricing concepts for broadband IP networks. *IEEE Communications Surveys*, 3(2), 2000.

[13] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. *Distributed Computing*, 18(1):61–72, 2005.

[14] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, 2001.

[15] R. Gibbens and F. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35(12):1969–1985, 1999.

[16] A. Gupta, D. Stahl, and A. Whinston. Priority pricing of integrated services networks. *Internet Economics*, pages 323–352, 1997.

[17] B. Hajek and S. Yang. Strategic buyers in a sum-bid game for flat networks. *Manuscript, http://tesla. csl. uiuc. edu/hajek/Papers/HajekYang. pdf*, 2004.

[18] L. He and J. Walrand. Pricing and revenue sharing strategies for internet service providers. *INFOCOM 2005. 24th*

Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, 1, 2005.

[19] T. Henderson, J. Crowcroft, and S. Bhatti. Congestion Pricing: Paying Your Way in Communication Networks. 2001.

[20] G. Huston. Interconnection, peering, and settlements. *Proc. INET*, 1999.

[21] R. Johari and J. Tsitsiklis. Routing and peering in a competitive Internet. *Prof. IEEE CDC*, 2, 2004.

[22] F. Kelly. Charging and accounting for bursty connections. *Internet Economics*, pages 253–278, 1997.

[23] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37, 1997.

[24] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[25] P. Laskowski and J. Chuang. Network monitors and contracting systems: competition and innovation. *ACM SIGCOMM Computer Communication Review*, 36(4):183–194, 2006.

[26] J. Le Boudec. Rate Adaptation, Congestion Control and Fairness: A Tutorial. *Web page, November*, 2005.

[27] J. Levy, H. Levy, and Y. Kahana. Top percentile network pricing and the economics of multi-homing. *Annals of Operations Research*, 146(1):153–167, 2006.

[28] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein. Internet Economics: The use of Shapley value for ISP settlement. *Proc. CoNEXT*, 2007.

[29] J. MacKie-Mason, L. Murphy, and J. Murphy. Responsive Pricing in the Internet. *Internet Economics*, pages 279–304, 1997.

[30] J. MacKie-Mason and H. Varian. Pricing the Internet. *Public Access to the Internet*, pages 269–314, 1995.

[31] J. Musacchio, S. G., and W. J. Network Neutrality and Provider Investment Incentives. *Proc. Network Economics*, 2007.

[32] W. Norton. Internet Service Providers and Peering. *Proceedings of NANOG*, 2001.

[33] A. Odlyzko. Paris metro pricing for the internet. *Proceedings of the 1st ACM conference on Electronic commerce*, pages 140–147, 1999.

[34] A. Odlyzko. Should Flat-Rate Internet Pricing Continue? 2000.

[35] M. Pauly. Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. *Quarterly Journal of Economics*, 88(1):44–62, 1974.

[36] M. C. F. Program. The broadband incentive problem. 2005.

[37] J. Roberts. Internet Traffic, QoS, and Pricing. *Proceedings of the IEEE*, 92(9):1389–1399, 2004.

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

[39] S. Shavell. On Moral Hazard and Insurance. *Quarterly Journal of Economics*, 93(4):541–562, 1979.

[40] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *Telecommunications Policy*, 20(3):183–201, 1996.

[41] J. von Neumann and O. Morgenstern. Theory of Games and Economic Behavior. *Princeton, NJ: Princeton*, 1947.

[42] M. Weiss and S. Shin. Internet Interconnection Economic Model and its Analysis: Peering and Settlement. *Communication Systems: The State of the Art*, 2002.

[43] E. Winter. The Shapley Value. *Handbook of Game Theory with Economic Applications*, 3:2025–2054, 2002.

[44] T. WU. Network Neutrality, Broadband Discrimination.

[45] S. Yang and B. Hajek. VCG-kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals. *IEEE Jounal of Selected Area in Communications, Special Issue on Non-Cooperative Behavior in Networking*, 25(6), 2007.