ISPs' Traffic Engineering and Peering Strategy

WANG, Hui

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

Abstract of thesis entitled:
ISPs' Traffic Engineering and Peering Strategy
Submitted by WANG, Hui
for the degree of Doctor of Philosophy
at The Chinese University of Hong Kong in September 2007

The Internet has quickly evolved into a vast global network owned and operated by thousands of interconnected Internet Service Providers. Each of these ISPs, as one autonomous system, has its individual economic interests. ISPs can achieve their objectives through *peering strategy* and *interdomain traffic engineering*. These two issues are important for ISPs' business and have significant implications on the Internet architecture.

Our study on interdomain traffic engineering focuses on AS Path Prepending (ASPP), a popular way for inbound traffic engineering. In order to improve the current situation that ISPs often practise this approach in a trial-and-error basis, we propose a greedy algorithm to help ISPs perform this approach systematically and efficiently. Then we demonstrate two fundamental issues of decentralized selfish traffic engineering, *routing instability* and *global network performance degradation*, based on an abstract model where ISPs perform traffic engineering for their individual load balance. We also present a real-world pathologic case of prepending instability from our measurement study. Some simple guidelines are given for ISPs to avoid such routing instability.

Over the past several years, numerous types of "overlay" networks change the interdomain traffic pattern and ISPs lose the

routing control of some interdomain traffic flows due to the application layer routing. As a result, some ISPs may provide unintended transit service for other local ISPs. It upsets the traditional business model and makes ISPs' peering strategies more complicated.

Our work on peering strategy is to help ISPs understand the economic implications of various traffic patterns and make proper decisions to optimize their business. We first conduct an economic analysis for an overlay streaming network to gain some insights on the *free ride* phenomenon. We further improve the analysis by taking the response of subscribers into consideration and formulate the dynamic market as a multi-leader-follower game to capture the *Nash Equilibrium* of the routing tussle among the major players of the Internet marketplace. Based on this framework together with a *gravity traffic model*, we present some important observations on the implications of overlays on ISPs' peering strategy.

論文題目：ISP 的流量工程和互連策略
院校與學系：香港中文大學信息工程系
學生：王會 **導師：邱達民 呂自成**

中文摘要

互聯網在近年來迅速发展成為由上仟個 ISP 共同運行的大型網絡。每一個 ISP 都是一個獨立的自治繫統，有自己的經濟利益。ISP 可以通過流量工程和合適的互連策略來優化自己的業務運作。互連策略是指如何選擇合適的 ISP 與之互連、選擇什么類型的互連關系以及多大的互連帶寬。如果一個 ISP 購買了多個綫路，它就可以進行所謂的域間流量工程，也就是為自己的進出流量選擇路由。互連策略和域間流量工程是 ISP 業務運作的重要方面，對互聯網體系結搆有着重要的影響。

本文通過對 AS Path Prepending (ASPP)的分析來研究流量工程中的各種問題。ASPP 是一種控製入流量的流量工程技術。盡管目前這種技術在互聯網上的應用已經比較廣泛，但 ISP 必須通過反復試驗才能夠取得合適的傚果。為了改善這種情況，我們在本文中提出一個貪婪算法來幫助 ISP 系統有效地使用這種技術。我們對 ISP 進行流量工程的行為進行抽象建模，並基于該模型分析了互聯網上分佈式的流量工程可能導緻的多種問題，例如路由振蕩和網絡整體性能惡化。根據我們對一段時間內 ISP 實施 ASPP 的行為進行的測量研究，我們在本文也給出了一個實際的路由振蕩的例子。最后我們提出了一些簡單的指導原則來幫助 ISP 避免 ASPP 引起的路由振蕩。

互聯網規模較小的時候互連策略是比較簡單的。隨着互聯網的发展，互連策略也變得越來越復雜。在過去幾年裏互聯網上出現了各種各樣的重叠網。這些重叠網改變了網間數據流的模式，通過應用層路由使 ISP 失去了對部分網間流量的路由控製，導緻 ISP 可能向其他 ISP 提供沒有包含在互連協議中的服務，且不能得到收益。這給傳統的業務模型和互連策略提出了新的挑戰。

我們的目的是為了幫助 ISP 理解各種網間流量模式對 ISP 業務的影響，使之能夠作出合適的決策來優化自己的業務。我們首先對一個媒體流重叠網進行分析，從而對“free ride”現象有初步的瞭解。在我們的路由模型中，該網絡的路由是為了優化重叠網內所有流量的傳輸性能，從而反映了傳統路由和重叠網路由的重要不同。然后我們進一步完善我們的理論分析框架。我們考慮到互聯網用戶會根據 ISP 服務質量的不同為自己選擇合適的服務商，因此我們將市場的動態變化建模成一個 multi-leader-follower 博弈，並對該博弈的納什平衡點進行分析，從而理解互聯網市場重要參與者之間的路由爭議所產生的經濟影響。基于這個博弈框架和一個通用的重力流量模型，我們分析了重疊網對 ISP 互連策略的一些重要影響。

Acknowledgement

Let me begin by expressing my gratitude to my supervisors, Professor Dah Ming Chiu and Professor John C.S. Lui, for inspiring me this interesting research topic, the constructive suggestions and comments, their amazing patience and their continuous encouragement throughout my study. I am lucky to have such nice supervisors, always willing to make the time to meet and discuss my work, and giving me complete freedom to deal with my research in the style I was most comfortable with. Their intellect, rigor, enthusiasm and humor will always inspire me.

A special thanks go to Professor Rocky Chang for his helpful discussion and thoughtful comments on my work on interdomain traffic engineering. I have troubled him with a lot of questions, and learnt a great deal from him.

I greatly appreciate the efforts of all the members in my committee for their time, their interest, and their feedback which helped shape this thesis.

My days as a graduate student have been wonderful. I must also thank my colleagues and officemates for helpful discussions and providing a friendly environment. I would like to thank all of the people that make ANSR a cheerful place to stay and make my life brighter for their presence.

To my parents, words cannot express my love and gratitude. They have loved me, supported me, always trusting me to make my own decisions, but have never expected, or asked for, anything from me.

Contents

Abstract **i**

Acknowledgement **iv**

1 Introduction **1**
1.1 Interconnection of ISPs 1
1.1.1 Peering Relationship 3
1.1.2 Competition and Cooperation 4
1.1.3 Technology and Business 5
1.2 Running an ISP 6
1.2.1 Peering Strategy 7
1.2.2 Traffic Engineering 8
1.3 Organization of the Thesis 10
1.4 Publications 13

2 Background Study **15**
2.1 Routing in the Internet: BGP 15
2.1.1 Policy-based Routing 19
2.1.2 Best Route Decision Process 21
2.1.3 A Discussion on BGP Routing System . . 22
2.2 Interdomain Traffic Engineering 23
2.2.1 Control of the Outgoing Traffic 25
2.2.2 Control of the Incoming Traffic 26
2.3 Routing Instability 29
2.3.1 Policy Induced Routing Instability 29

2.3.2 Solutions to Routing Instability 31
2.4 Overlay Networks 33
2.4.1 CDN and Routing Overlay 34
2.4.2 Impact of Overlays 35

3 Interdomain Traffic Engineering 38
3.1 Popularity and Justifications 40
3.2 Terminologies and Notations 41
3.3 Greedy ASPP Search Algorithm 44
3.3.1 A Greedy ASPP Search Algorithm 44
3.3.2 Example: Single AS Load Balancing . . . 47
3.3.3 Performance Analysis of the Algorithm . . 50
3.4 ASPP by Multiple ASes 53
3.4.1 ASPP for Load-independent Reasons . . . 54
3.4.2 ASPP for Load-dependent Reasons 56
3.5 Finding ASPP Instability in the Internet 65
3.5.1 Finding Potential Conflicts 66
3.5.2 A Pathological Example 68
3.6 Guidelines to Avoid ASPP Instability 71
3.7 Related Work 74
3.8 Summary 77

4 Economic Analysis of Overlay Streaming 80
4.1 The Market under Study 83
4.2 Methodology for Economic Analysis 85
4.3 Traffic and Routing 86
4.3.1 Routing Model 87
4.3.2 Analysis of the Routing Solution 90
4.4 Economic Analysis of Different Scenarios 92
4.4.1 No Peering Scenario 92
4.4.2 Unlimited Capacity Peering Scenario . . . 94
4.4.3 Limited Capacity Peering Scenario 99
4.5 Numerical Example 101

4.5.1 Private Link Provisioning 101
4.5.2 Peering Link Provisioning 102
4.6 Markets with Multiple ISPs 103
4.6.1 Unlimited Capacity Peering Scenario . . . 104
4.6.2 Limited Capacity Peering Scenario 106
4.7 Summary 107

5 Inter-ISP Traffic and Peering Strategy 109
5.1 Business Model 112
5.2 Game Theoretic Framework 114
5.3 Solution of the Game 118
5.4 Inter-ISP Traffic Model 121
5.4.1 Gravity Modeling Framework 122
5.4.2 Interpretation of Performance Factor . . . 125
5.4.3 Interpretation of Repulsive Factor 126
5.4.4 Inter-ISP Traffic Function 128
5.5 Case Studies of ISPs Peering Game 129
5.5.1 Provisioning Strategy: Web Traffic 130
5.5.2 Provisioning Strategy: Overlay Traffic . . 133
5.5.3 Provisioning Strategy: Traffic Transition . 137
5.5.4 Peering Strategy 138
5.5.5 Discussion on Subscribers' Choices 140
5.6 Related Work 143
5.7 Summary 147

6 Conclusion 149

A Solving the Optimization Problem 154
A.1 No Peering: $y_1^* = 0$ and $y_2^* = 0$ 155
A.2 Unlimited Capacity Peering: $u_9 \neq 0$ 156
A.2.1 Condition for Unlimited Capacity Peering 157
A.2.2 Solution for Unlimited Capacity Peering . 161
A.3 Limited Capacity Peering: $u_9 = 0$ 162

Bibliography **163**

List of Figures

2.1 An Example of Network with Different Relationships 16
2.2 The BGP System BAD GADGET 30

3.1 Example Network for Studying Prepending Algorithm 47
3.2 Comparison of the Result from the Greedy Algorithm with the Optimal Result 52
3.3 The Distribution Function of Iteration Numbers . 53
3.4 Screen Shot of the GUI of Simulation Tool. 60
3.5 Example Network to Study Routing Convergence 62
3.6 Routes to Prefix $\mathcal{P}$ from Route Views 70

4.1 Example to Illustrate Routing Control. 81
4.2 The Local Market under Study. 84
4.3 Routing Model. 88
4.4 Feasible Region. 90
4.5 Different Outcomes of Unlimited Capacity Peering. 97
4.6 Break-even Price of Two ISPs as We Vary c_1^o. . . 102
4.7 Break-even Price of Two ISPs as We Vary c_2^p. . . 103

5.1 The Local Market under Study. 113
5.2 Overview of the Game Theoretic Framework. . . . 117
5.3 Overview of the Traffic Model. 122
5.4 α_1^* in Different Scenarios (Web Traffic). 131
5.5 p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Web Traffic). . . 132
5.6 Performance Sensitivity Function $G(\delta)$. 134

5.7 α_1^* in Different Scenarios (Overlay Traffic). 135
5.8 p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Overlay Traffic). 136
5.9 α_1^* in Different Scenarios (Peering). 138
5.10 p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Peering). 139
5.11 α_1^* under Different Peering Capacities. 140
5.12 p_i^* and $\mathcal{R}_i^*$ under Different Peering Capacities. . . 141
5.13 Some Subscribers Do Not Choose Internet Access Service. 142

A.1 Conditions for No Peering Scenario 157
A.2 Unlimited Capacity Peering Scenario: Case 1. . . 158
A.3 Unlimited Capacity Peering Scenario: Case 2. . . 159
A.4 Location of Intersection Points 160

List of Tables

3.1 AS_1 Executing the Greedy ASPP Search Algorithm for Load Balance. 49
3.2 The Network Routing Converges. 63
3.3 Routing Instability Caused by Interference of ASPP. 64
3.4 Convergence to Worse Global State. 65
3.5 Prepending Changes in Routes to Prefix $\mathcal{P}$ 69

Chapter 1

Introduction

Summary

The Internet is a "network of networks" operated by many ISPs. These ISPs interconnect with each other at IXPs under different business relationships. In this chapter, we briefly describe the two primary peering relationships, the cooperation and competition among ISPs, and the tied destiny of the business model and technology model within the Internet architecture. We also introduce the two important tasks in running one ISP, *i.e.*, *peering strategy* and *traffic engineering*. Finally, we present the organization of this thesis.

1.1 Interconnection of ISPs

The Internet is a worldwide, publicly accessible network of interconnected computer networks that transmit data by packet switching using the standard Internet Protocol (IP). It is a "network of networks" that consists of millions of smaller domestic, academic, business, and government networks, which together carry various information and services, such as electronic mail, online chat, file transfer, and the interlinked web pages and other documents of the world wide web [2].

The Internet is operated by many Internet Service Providers (ISPs). An ISP (also called Internet Access Provider or IAP) is a business or organization that provides to consumers access to the Internet or other related services, such as Internet transit, domain name registration and hosting *etc*. In the past, most ISPs were run by the phone companies. Now, ISPs can be started by just about any individual or group with sufficient money and expertise.

ISPs employ a range of technologies to enable consumers (the end-users of the Internet) to connect to their networks, such as dial-up, DSL (typically ADSL), Ethernet, Metro Ethernet, and Gigabit Ethernet. With the increasing popularity of downloading music and online video and the general demand for faster page downloads, higher bandwidth connections are becoming more popular.

End-users generally want to have access to all other possible end-users, regardless of the network they belong to. To provide such universal connectivity to their users, ISPs must interconnect with each other, forming one large, global entity to share their network infrastructure. The point where two or more ISP networks meet is generally called an Internet eXchange Point (IXP). An IXP is where one ISP advertises routes (destination networks) to another, and where data packets flow from one ISP to another. Logically, an IXP consists of routers interconnected through a variety of layer-2 (Ethernet, FDDI, or ATM, for example) and layer-3 (IP routers) mechanisms. A public IXP is owned and operated by a third party and is open to any ISP that wishes to interconnect with other ISPs there. A private IXP is a direct point-to-point interconnection between ISPs [10].

1.1.1 Peering Relationship

The business relationship formed by the parties interconnecting at an IXP can take different forms. The major relationship is *transit* relationship, where a smaller ISP (in terms of size, number of customers, geographic reach, and so on) pays a transit fee to a large ISP (upstream ISP) for Internet access, just as the customers of the small ISP pay it for Internet access. Transit relationship is also called as "*provider-to-customer*" relationship. In the simplest case, a single connection is established to an upstream ISP (provider), and the ISP (customer) uses this connection to send or receive any data to or from parts of the Internet beyond its own network; in turn, the upstream ISP uses its own upstream connection, or connections to its other customers (usually other ISPs) to allow the data to travel from source to destination. In reality, the situation is often more complicated. For example, ISPs with more than one Point of Presence (PoP) may have separate connections to an upstream ISP at multiple PoPs, or they may be customers of multiple upstream ISPs and have connections to each one at one or more of their PoPs.

ISPs may also engage in *free peering* relationship, where multiple ISPs interconnect with one another at a peering point or Internet exchange point, allowing the routing of data between their networks, without charging one another for that data flows. The data flows would otherwise have passed through their upstream ISPs, incurring charges from the upstream ISPs. This kind of relationship is also called *peer-to-peer* or *peering* for short.

From the viewpoint of business relationship, we can define two types of ISPs, namely, *transit ISP* and *stub ISP*. Stub ISPs are the ISPs that have no customer (or client ISP), while transit ISPs are ISPs that have customers. Transit ISP provides Internet connectivity to other ISPs' networks by forwarding all types of traffic across its network. Stub ISP, on the other hand, does not provide transit service for other ISPs and only sends or receives the traffic of its own network. Moreover, ISPs who require no upstream, and have only customers and/or peers, are called *Tier 1* ISPs, indicating their status as ISPs at the top of the Internet hierarchy. By definition, a Tier 1 network does not purchase IP transit from any other network to reach any other portion of the Internet. Therefore, in order to be a Tier 1, a network must peer with every other Tier 1 network.

Free peering might seem ideal for an ISP in terms of cost and performance, but the customer routes acquired from the peering partners may not be able to provide access to the entire Internet. A transit arrangement might be expensive, and packets might incur extra latency crossing the transit provider's network, but it assures access to all routes in the global routing table. In general, more ISPs are moving toward peering relationships to help lower costs and improve performance by minimizing the traffic flowing over the links purchased from provider ISPs. This trend is being facilitated by the introduction of more IXPs, particularly those with gigabit Ethernet (GE) and optical technologies, which enable higher-performance switching capacity.

1.1.2 Competition and Cooperation

The Internet is a collection of interconnected component networks that share a common addressing structure, a common view of routing and traffic flow, and a common view of a naming system. This interconnection environment spans a highly diverse set of more than 50,000 component networks, and this number continues to grow and grow.

One of the significant aspects of this environment is the competitive Internet service industry, where many thousands of enterprises, both small and large, compete for market share at a regional, national and international level. Underneath the veneer of a highly competitive Internet service market is a somewhat different environment, in which every ISP network must interoperate with neighboring Internet networks in order to produce a delivered service outcome of comprehensive connectivity and end-to-end service. No ISP can operate in complete isolation from others while still participating in offering public Internet services, and therefore, every ISP must not only coexist with other ISPs but also must operate in cooperation with other ISPs.

As a result, ISPs compete with each other on price, performance and reliability *etc.* at a level of business, but they also must cooperate with each other to provide global connectivity to all other attachments on the Internet at a level of technology.

1.1.3 Technology and Business

Technology and business models share a common evolution within the Internet. The business evolution is always fueled by technology. To enable deployment of the technology within a service environment there is also the need to create a robust and stable business model. An example is that while IP multicast offers by far the most efficient delivery vehicle for large-scale multiparty communications, few service providers deploy it due to the lack of corresponding business model, choosing instead to consume bandwidth and host resources with multiple point-to-point connections.

This tied destiny of technology and business factors is perhaps most apparent within the area of the interconnection of ISPs. We can look closely at the interaction between the capabilities of the technical protocols, their translation into engineering deployment and the consequent business imperatives. For example, the current interdomain routing protocol allows ISPs to express their *routing policies*, so ISPs can filter out the routes they do not want, therefore ISPs can control the routing signalling and traffic flows at network layer. Based on this control, ISPs can set up peering agreements according to the routes exchange. However, this situation might be changed after the emergence of overlay networks. Technically, ISPs cannot control the traffic flows at application layer, so the routes exchange of ISPs at network layer is not enough to reflect the service they provide each other. As we will demonstrate in this thesis, this new phenomenon upsets the traditional business model and makes the negotiation of benefit and cost in the interconnection more difficult.

In this thesis we will examine both the technical and business aspects that surround the ISP interconnection and their interaction, including some issues from the technical architectures of such environments and the business motivation for the interconnection structure.

1.2 Running an ISP

Each ISP in the Internet has its individual economic interests. ISPs are driven to lower costs, optimize performance, and maximize revenue. When network operators run an ISP, they need to undertake two important tasks:

- peering strategy: determine where and with whom to peer or transit, and the peering capacity;
- traffic engineering: determine the incoming link or the outgoing link for all traffic across the ISP border.

The peering strategy and the interdomain traffic engineering are very important for ISPs to run their business and have significant influence on the Internet architecture evolvement. The following is a brief description of these two aspects of ISPs' operation.

1.2.1 Peering Strategy

One ISP needs to determine with which ISPs it connects, what kind of peering agreements it makes with these connected ISPs, and how much bandwidth the connections should be. If the ISP appears at multiple geographical locations, it further needs to determine the peering locations.

The most frequent relationship between ISPs is the provider-to-customer relationship and it is also the origin of most of the interdomain cost of an ISP. A stub ISP usually tries to maintain at least two links from its providers for performance and redundancy reasons. In addition, large ISPs typically try to obtain peer-to-peer relationships with other ISPs and then share the cost of the link with other ISPs. Negotiating the establishment of those peer-to-peer relationships is often a complicated process since there are a lot of technical and economical factors need to be taken into account [57] [55] [58]. We list some of the advantages and obstacles of the free peering relationship here.

Advantages of Free Peering

Free peering is one of the most important ways for ISPs to improve the efficiency of operation. ISPs seek peering relationships primarily for two reasons. First, peering decreases the cost and reliance on purchased Internet transit. Second, peering lowers inter-ISP traffic latency. By avoiding a transit provider hop in between ISPs, traffic between peering ISPs has lower latency. The second motivation is especially important under usage-based traffic billing scenario. Since loss and latency slows traffic consumption, the usage-based billing ISPs benefit from a lower latency, lower packet loss Internet. In these years, the rapidly declining transit prices and relatively static exchange point prices has caused a shift in the motivation for peering from cost savings to performance improvements.

For the Tier 1 ISPs, since they do not pay for transit, their motivation of free peering is to provide sufficient interconnection bandwidth to support their customer base and their growth. Generally, Tier-1 ISPs only want to peer with Tier-1 ISPs.

Obstacles of Free Peering

However, there are also some obstacles for peering. First, as a "peer" there are no Service Level Agreements (SLAs) to guarantee rapid repair of problems. Some service providers prefer a transit (customer) relationship with ISPs for business reasons, arguing that the threat of lost revenue is greater than the threat of terminating a peering arrangement if performance of the interconnection agreement is inadequate. Second, the peer-to-peer agreement often requires traffic symmetry on the link, but sometimes this requirement cannot be satisfied. Third, free peering consumes resources (router interface slots, circuits, staff time, *etc.*) that could otherwise be applied to revenue generation. Note there might be the potential for transit sales if not peer. The last point, for tier 1 ISPs, they compete on the basis of better performance, so they may want to avoid free peering with other ISPs because they do not want to improve the performance of its "peers" and make them a more powerful competitor.

In these few years, the emergence of overlay networks brings new challenges to ISPs' peering strategy and makes the interconnection negotiation more complicated.

1.2.2 Traffic Engineering

If one ISP decides to connect with other ISPs through more than one links, it can further optimize the way traffic enters or leaves its network based on its business and performance considerations, which is the task of *interdomain traffic engineering.* Optimizing the way traffic enters or leaves a network means to favor one link over another to reach a given destination or to receive traffic from a given source. This type of interdomain traffic engineering can be performed by tweaking the BGP routes of the ISP.

Content providers that host a lot of web or streaming servers and usually have several customer-to-provider relationships with transit ISPs will try to optimize the way traffic leaves their networks. On the contrary, access-providers that serve small and medium enterprises, dialup or ADSL users typically wish to optimize how Internet traffic enters their networks. In addition, a transit ISP will try to balance the traffic on the multiple links it has with its peers, to improve the network resource utilization factor.

The state of the art for interdomain traffic engineering is extremely primitive. The IETF's Traffic Engineering Working Group, which has focused almost exclusively on intradomain traffic engineering, recently noted that interdomain traffic engineering "is usually applied in a trial-and-error fashion. A systematic approach for interdomain traffic engineering is yet to be devised". Operators make manual changes in the routing policies without a good understanding of the effects on the flow of traffic or the impact on other domains [32]. Ultimately, this *ad hoc* approach to interdomain traffic engineering must evolve into mature, well-tested guidelines and mechanisms. Recent previous work has presented a high-level overview of interdomain traffic engineering and characterized the traffic data which must be measured to perform interdomain traffic engineering. In addition, there have been several commercial products that help large campus and corporate networks balance load over connections to multiple upstream providers; however, the automated traffic engineering may bring new problems to the Internet, which is a large scale distributed system.

1.3 Organization of the Thesis

In this thesis we focus on the two primary aspects of ISPs' operation: interdomain traffic engineering and peering strategy. Chapter 2 presents a brief introduction of preliminary background information. We introduce the current interdomain routing protocol and popular methods to implement interdomain traffic engineering. We also describe the famous routing instability problem of Border Gateway Protocol, which is induced by policy disputes, and some possible solutions to this problem. As a supplement of network layer routing controlled by ISPs, overlay networks appeared in these years to favor applications' needs through application layer routing. In the last section of this chapter, we introduce the concept of overlay networks and also give a brief overview of their impacts.

Chapter 3 focuses on various aspects of interdomain traffic engineering. ISPs perform traffic engineering based on local objectives and local information. Such local decisions may potentially result in global network performance degradation or routing instability issues. In this chapter, we restrict our attention to the *AS Path Prepending* approach to illustrate the fundamental issues of decentralized selfish traffic engineering in the Internet. Although this approach has been extensively practised by many ISPs, there still lacks a systematic study of this approach and the basic understanding of its effectiveness, therefore ISPs often practise this method in a trial-and-error basis. We first propose a greedy algorithm to help ISPs to practise this approach systematically and efficiently. Then we demonstrate that how automated traffic engineering or policy conflict introduces routing instability problem. We also evaluate the global effect of the local traffic engineering practices based on our performance metrics. Finally, from our measurement study of the dynamic prepending behavior of ISPs in the real Internet, we show a real-world pathologic case of prepending instability. Some guidelines are given as the first step to study the method to avoid such routing instability problem.

Over the past several years, we have seen the emergence of numerous types of "overlay" networks in the Internet to enhance or modify the basic functioning of traffic handling within the Internet. The overlays change the pattern of inter-ISP traffic by application layer routing, so they have a clear technical impact on the Internet. Since the inter-ISP traffic pattern also determines money flows between providers, overlays also influence the commercial relationships on the Internet, which in turn, gives rise to policy implications. For example, the traffic routed at application layer is not visible and controllable by policy-based interdomain routing, and may cause some ISPs to provide unintended transit service for other local ISPs. Chapter 4 and Chapter 5 focus on the economic implications of overlay networks on ISPs' peering strategy.

In Chapter 4 we conduct an economic analysis for an overlay streaming network to gain some insights on the *free ride* phenomenon brought by overlay networks. To capture the important difference between how overlay traffic and traditional web traffic are routed in the Internet, we model the routing in the network as to optimize the performance of transferring each byte of its subscribers' demand. The overlay streaming traffic is routed at the application layer according to the availability and performance of all inter-ISP paths, oblivious of policies configured in interdomain routing. In this analysis, we deploy *break-even price* as the metric to evaluate the economic efficiency of one ISP. The analysis shows that ISPs must control their provider link capacity as well as free-peering link capacity they provide to their peering ISPs to maintain certain parity between different incoming link capacity among all ISPs. Only when such parity is maintained, ISPs can sustain their relative economic positions hence economic equilibrium despite of application layer routing. It reveals the fact that the application layer routing of overlay networks upsets ISPs' traditional peering strategy.

Chapter 4 focuses on a particular kind of overlay network, therefore it cannot be applied directly on other overlay traffic types. Moreover, the analysis is conducted in a static market, where both the market share distribution and the provisioning of all links are fixed. As a result, that economic analysis does not take the response of subscribers into consideration. In addition, the evaluation based on break-even price only reveals the influence direction of overlay traffic at a particular point of the market, but cannot capture the stable state (equilibrium) of a dynamic market.

In Chapter 5, we improve the analysis by taking the response of subscribers into consideration - subscribers can shift to other ISPs that provide better service, which is what happens in the real Internet marketplace. The dynamic market of two ISPs is formulated as a multi-leader-follower game to study the output of the routing tussle between applications and ISPs, and capture the decisions of the major players in the Internet marketplace - the subscribers (users), and the ISPs playing different roles in providing transit service. We also propose a gravity traffic model for the inter-ISP traffic flows simultaneously controlled by ISP peering and provisioning (of its link capacities), and network layer and application layer routing. Based on the game theoretic formulation together with the traffic model, we derive some insights on peering strategies in scenarios with overlay traffic. This framework can also be used by ISPs to understand various traffic patterns, evaluate their peering strategies and make proper decisions to optimize their business.

1.4 Publications

The work presented in this thesis is the result of original research conducted by the author. Parts of it have been published, or submitted for publication, as follows:

Chapter 3:

- Jessie Hui Wang, Rocky K.C. Chang, Dah Ming Chiu, and John C.S. Lui, "Characterizing the Performance and Stability Issues of the AS Path Prepending Method: Taxonomy, Measurement Study and Analysis", in *Proceedings of ACM SIGCOMM ASIA Workshop 2005*, Beijing, China, April 2005.

- Jessie Hui Wang, Rocky K.C. Chang, Dah Ming Chiu, and John C.S. Lui, "Inter-AS Inbound Traffic Engineering via ASPP", *IEEE Transactions on Network and Service Management*, *4(1)*, June 2007.

Chapter 4 and 5:

- Jessie Hui Wang, Dah Ming Chiu, and John C.S. Lui, "Modeling the Peering and Routing Tussle between ISPs and P2P Applications", in *Proceedings of Fourteenth IEEE IWQoS*, Yale University, New Haven, CT, USA, June 2006.

- Jessie Hui Wang, Dah Ming Chiu and John C.S. Lui, "A Game-theoretic Analysis of the Implications of Overlay Network Traffic on ISPs' Peering", submitted to *Computer Networks* (*Elsevier*).

Others

- Yangfan Zhou, Michael R. Lyu, Jiangchuan Liu, and Jessie Hui Wang, "PORT: A Price-Oriented Reliable Transport

Protocol for Wireless Sensor Networks", In *Proceedings of the 16th IEEE International Symposium on Software Reliability Engineering (ISSRE'05)*, Chicago, IL, USA, November 2005.

□ **End of chapter.**

Chapter 2

Background Study

Summary

This chapter presents a brief introduction of preliminary background information. We introduce the current interdomain routing protocol and popular methods to implement interdomain traffic engineering. We also describe the famous routing instability problem of Border Gateway Protocol, which is induced by policy disputes, and some possible solutions to this problem. Finally, we introduce the concept of overlay network, which provides the application layer routing as a supplement of the network layer routing controlled by ISPs. We also give a brief overview of their impacts.

2.1 Routing in the Internet: BGP

Routing in the Internet is based on the concept of "autonomous systems" (AS). AS is the unit of routing policy in the modern world of exterior routing, and are specifically applicable to protocols like EGP (Exterior Gateway Protocol, now at historical status) and BGP (Border Gateway Protocol). In RFC 1930, an

AS is defined as a connected group of one or more IP prefixes run by one or more network operators which has a *single* and *clearly defined* routing policy. Typically, an AS is a collection of IP networks and routers under the control of one *single* entity (typically an ISP). Sometimes one AS can be run by multiple entities, but it must present a common routing policy to the Internet. In addition, one ISP may run multiple ASes due to historical reasons or technical reasons. In this thesis, we use the two terminologies "AS" and "ISP" interchangeably. When we talk about technical stuffs, we use "AS" more frequently, while "ISP" is often used to emphasize the business behavior of one entity.

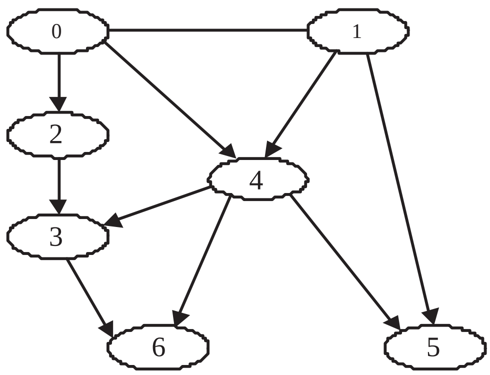

Figure 2.1: An Example of Network with Different Relationships

It is customary to use a connected graph $G = (V, E, B)$ to represent the interdomain network topology and business relationship. Each node $v \in V$ represents an entire AS; and each edge $e \in E$ denotes a logical link connecting two ASes. For each edge e, $B(e)$ defines the business relationship between the two ASes for which e connects. Figure 2.1 illustrates a network of seven ASes with edges that convey different relationships. In particular, a provider-to-customer relationship is represented by a directed edge wherein the pointed node represents the customer, whereas a free peering relationship between two ASes is represented by a undirected edge. Therefore, in this figure AS_4 is a provider for AS_3, AS_5 and AS_6 while AS_2 and AS_4 are customers of AS_0. There is one free peering relationship in the figure and it is between AS_0 and AS_1.

Routing in the Internet is performed by the interplay of interdomain routing protocol and intradomain routing protocol. Routers on the border of an AS exchange reachability information using interdomain routing protocol, which is responsible for determining the AS-level forwarding path. Inside each AS, intradomain routing protocol determines shortest paths between every router within the network. Interdomain routing protocol also determines the set of best egress points for each destination prefix. When there are multiple equally good egress points according to interdomain protocol, intradomain protocol provides a ranking of these egress points for a given ingress point in terms of closeness.

BGP is the only *de facto* interdomain routing protocol in today's Internet [19]. The purpose of BGP is to allow two different ASes to exchange routing information so that data traffic can be forwarded across the AS border. For scalability reasons, the interdomain routing protocol is only aware of the interconnections between distinct domains, it does not know any information about the content of each domain.

BGP is a protocol based on the distance vector algorithm [37] and it uses TCP as its transport protocol. After a BGP session is established, each end advertises a route for every prefix that it wants the other end to know. After these initial messages have been exchanged, a node only needs to inform the node at the other end about any route changes.

Routing information shared among peers in BGP has two forms: announcements and withdrawals. A route announcement indicates that a router has either learned of a new network attachment or has made a policy decision to prefer another route to a network destination. Route withdrawals are sent when a network is no longer reachable. We distinguish between explicit and implicit withdrawals. Explicit withdrawals are those associated with a withdrawal message; whereas an implicit withdrawal occurs when an existing route is replaced by the announcement of a new route to the destination prefix without an intervening withdrawal message. In order to reduce the packet head overhead, a BGP update message may contain multiple route announcements and withdrawals.

When a BGP router advertises a prefix to one of its BGP neighbors, a number of attributes are associated with that announcement. Attributes are used heavily in BGP to carry a wide range of information. For instance, *AS-PATH*, which is an important attribute of BGP, contains the ASes through which the announcement for the prefix has passed. As an announcement is passed between ASes, each AS adds its AS number (ASN) to the AS-PATH attribute. This, by itself, is useful for the operators of the ASes to learn all the information of this route. However, it also provides the critical feature of detecting and preventing looping announcements. The *next hop* attribute indicates the IP address of the router associated with next hop along the path to the destination. One AS can advertise routes with different values of *multiple exit discriminator* (MED) attribute at different interconnection points between the two ASes, to encourage the recipient to select a particular exit point for sending traffic to the neighboring AS. *Local preference* is an locally assigned attribute to indicate the ranking of the paths learned from different routers in the AS. This attribute would not be announced to other ASes with the advertisement.

2.1.1 Policy-based Routing

BGP routing depends heavily on locally-configured policies. BGP has two kinds of routing policies: *import routing policy* and *export routing policy* (also referred to as *import filtering* and *export filtering*). Import policy determines which routes should be accepted from a neighbor and the preference with which those routes should be treated, while export policy determines which routes should be advertised to a neighbor.

A BGP router may receive multiple routes for the same destination prefix. Upon receiving a route advertisement, the router applies import policies to filter unwanted routes or to alter the attributes associated with the route. Network operators can configure import policies to influence path selection. Ultimately, the router invokes a decision process (see subsection 2.1.2) to select exactly one "best" route for each destination prefix among all the routes it receives. The router uses export policies to limit the distribution of routes to certain neighboring ASes, based on the business relationship between the two ISPs.

If an ISP accepts a route from a neighbor, it means this ISP agrees to provide transit service for the traffic destined to the prefix of this route, and the local preference can be used to do outbound traffic engineering. If an ISP advertises a route to one of its neighbors, it means this ISP would like to accept traffic destined to the prefix of this route from this neighbor, which is related with inbound traffic engineering.

The business relationships among ISPs play a crucial role in shaping the structure of the Internet and the end-to-end performance characteristics [26]. This can be illustrated by the following example. In Figure 2.1, AS_3 and AS_4 are providers of AS_6, which implies AS_6 pays for the traffic going through the link (AS_3, AS_6) and the link (AS_4, AS_6). Imagine that AS_3 wants to send traffic to AS_4. AS_6, being a customer to both AS_3 and AS_4, obviously does not want to provide transit service for its providers.

Routing policy used in combination with the BGP decision process is the key mechanism that BGP provides for ISPs to enforce business agreements made between two or more parties. In the above example, in order to avoid providing transit service between AS_3 and AS_4, AS_6's export routing policy will not allow it to announce the reachability information of AS_3 (AS_4) to AS_4 (AS_3). Obviously, an AS's business agreements with its neighbors determine how the AS routes the traffic that enters or exits its infrastructure. For this reason, we can say that AS's business agreements determine its routing policy.

In [12], the authors characterize the popular policies adopted by ASes in today's Internet as: (a) the *typical local preference import policy* and (b) the *selective announcement export policy*. Under the typical local preference policy, an AS prefers to use a customer link than a peering link to forward a packet, and it prefers to use a peering link than a provider link to forward a packet, provided that these links can reach the destination AS. This is natural since an AS can get revenue from the traffic going through its customer link, and it does not need to pay for the traffic going through its peering link. However it must pay for the traffic going through its provider link. Under the selective announcement export policy, an AS would not announce the routes learned from its providers or peers to other providers and peers, thus an AS does not provide transit service between its providers or its peers.

To illustrate, let us assume all ASes in Figure 2.1 obey typical local preference and selective announcement policy. Then routes with AS path (AS_5, AS_4, AS_6) or $(AS_5, AS_1, AS_4, AS_3, AS_6)$ may appear in the network, while routes with AS path (AS_1, AS_0, AS_4, AS_6) would not appear in this network, because AS_1 would select AS_4, instead of AS_0 as the next hop to reach AS_6 according to the typical local preference. Also, route with AS path (AS_1, AS_4, AS_0) would not appear since AS_4 would not announce AS path (AS_4, AS_0) to AS_1 according to the selective export policy.

2.1.2 Best Route Decision Process

A BGP router may learn multiple paths to the same destination prefix from its neighbors, it then select a single best path based on the attributes in those BGP advertisements. Although the complete details of the decision process are not part of the protocol specification, most router vendors adhere to a *de facto* standard. After certain routes are removed from consideration (*e.g.*, because they have a loop in the AS path, have an unreachable next hop, or were filtered by the import policy), the router applies a sequence of steps to narrow the set of candidate routes to a single choice. We list the first several steps as follows:

- 1. Highest local preference: Prefer routes with the highest local preference, where local preference is assigned by the import policy and is conveyed only via iBGP.
- 2. Shortest AS path: Prefer routes with the shortest AS path length, as conveyed in the BGP advertisement.
- 3. Lowest origin type: Prefer routes with the lowest origin type (IGP is preferable to EGP which is preferable to INCOMPLETE), as conveyed in the BGP advertisement or reset by the import policy.
- 4. Lowest MED: For routes with the same next-hop AS, prefer routes with the smallest MED value, as conveyed in the BGP advertisement or reset by the import policy.

Router vendors typically provide configuration options to disable one or more of these steps. Researchers may consider some (or all) of these steps to construct various models to capture the dynamic behavior of BGP at an abstract level [30] [49] [51].

2.1.3 A Discussion on BGP Routing System

BGP allows each ISP to express its policies for accepting, forwarding, and passing off packets using a variety of approaches to achieve different concerns. BGP then performs a distributed computation to determine the "best" path along which packets from each source to each destination should be forwarded. The resultant path taken by data packets is a result of individual decisions taken by each of the many ISPs that combine to form the Internet.

There are some problems in this routing formulation. The fundamental one is that the notion of "best" is not clear and it is in fact insufficient to fully express the routing task. "Best" is a single dimensional concept, but routing is a multi-dimensional problem. Individual ISPs, in making their routing decisions, may choose to optimize a wide variety of properties such as the cost of passing on a packet (to minimize the cost), load balance among different physical links (to maximize utilization and minimize congestion), bandwidth available to the traffic or transmission delay (to optimize performance). Further, because the management of each ISP chooses its own objective, different ISPs may choose to optimize different quantities, leading to an overall path that captures no simple notion of "best", and rarely if ever is best for the user.

We can make two broad statements about the present IP routing system. First, the route used for data is determined entirely by the ISPs, without input or control from end users or applications. Second, what is optimized by the routing system is an imprecisely defined mix of cost and ISP operational efficiency, rather than any metric directly related to application performance.

2.2 Interdomain Traffic Engineering

ASes that have more than one provider are called *multihomed* ASes. Motivated by the need to improve network resilience and performance, there is an increasing number of enterprises and campus networks connecting to the Internet via multiple providers. Network operators of these multihomed ASes must have control over the flow of traffic into, out of, and across their networks, which is called *interdomain traffic engineering*.

The goals of interdomain traffic engineering are diverse, such as to balance the traffic on all interdomain links of the AS, or to reduce the cost of carrying traffic on these links, or to improve the performance of the access service the AS provides [11]. Particularly, we list the following four example scenarios where traffic engineering is beneficial and necessary for the ISP.

- traffic fluctuation: Network failures and traffic fluctuations degrade user performance and lead to inefficient use of network resources. Network operators should adapt to the changes in the traffic distribution by adjusting the configurations of the routing protocols running on their routers.

- congested edge link: The links between domains are common points of congestion in the Internet. Upon detecting an overloaded edge link, an operator can change the interdomain paths to direct some of the traffic to a less congested link.

- upgraded link capacity: Operators of large IP backbones frequently install new, higher-bandwidth links between domains. Exploiting the additional capacity may require routing changes that divert traffic travelling via other edge links to the new link.

- violation of peering agreement: An AS pair may have a business arrangement that restricts the amount of traffic they exchange; for example, the outbound and inbound traffic may have to stay within a fixed factor. If this ratio is exceeded, an AS may need to direct some traffic to a different neighbor.

BGP is a flexible interdomain routing protocol that scales to the large number of ASes, and it enables network operators to achieve certain traffic engineering goals by tuning the local routing policies that affect the selection of the best path for a destination prefix. However, BGP was not designed with traffic engineering in mind, so it does not include any direct support for common traffic engineering tasks, such as balancing load across multiple links to a neighboring AS or directing traffic to a different neighbor.

Under this situation, solving the interdomain traffic engineering problems is difficult because of two reasons. First, the attributes available in BGP advertisements, the restrictions in the BGP decision process, and the constraints imposed by configuration languages all limit an operator's ability to tune routing policies to the prevailing traffic patterns. Second, the number of possible changes to routing policies is too large to exhaustively test all possibilities, some changes in routing policy can have an unpredictable effect on the flow of traffic. In the real world, finding out the appropriate policy configuration is difficult, and the result depends on many factors, such as the IGP parameters, the BGP advertisements received from neighboring ASes, the network topology, and the current traffic patterns *etc.*.

Although network operators cannot perform traffic engineering conveniently and directly based on BGP, there are some indirect BGP-based traffic engineering approaches as introduced in the following subsections [8].

2.2.1 Control of the Outgoing Traffic

To control how the traffic leaves its network, an AS must be able to choose which route will be used to reach a particular destination via its peering ISPs through the control of the decision process on its BGP routers. Two techniques are frequently used for outbound traffic engineering.

A first technique is to rely on the optional attribute "local preference", which is the first criteria of the BGP decision process. The following is an example of this technique. Consider a stub AS with two links toward one upstream provider: a link with high bandwidth and a link with low bandwidth. In this case, the BGP router of this AS could be configured to insert a low local preference to routes learned via the low bandwidth link and a higher value to routes learned via the high bandwidth link. A similar situation can occur for a stub AS connected to a cheap and a more expensive upstream provider.

A second technique, often used by large transit ISPs, is to rely on the tuning of intradomain routing parameters to influence how a packet crosses the transit ISP. The BGP decision process will select the nearest IGP neighbor when comparing several equivalent routes received via iBGP. Therefore, a tuning of the weights of its intradomain routing protocol will indirectly influence its outgoing traffic.

The control of the outgoing traffic using BGP is based on the selection of the best route among all available routes. This selection can be performed on the basis of various parameters, but it is limited by the diversity of routes received from upstream providers, which is determined by the connectivity and the policies of those ASes.

2.2.2 Control of the Incoming Traffic

The control of incoming traffic is based on a careful tuning of advertisements sent by the AS. Inbound traffic engineering is generally thought as more difficult than outbound traffic engineering because an AS cannot control the routing decisions of other ASes directly [6].

The first method is to divide the AS's address space in distinct prefixes and announce different route advertisements on different links. However, this *selective advertisements* method suffers from an important drawback: if a link fails, the prefixes that were announced on the failed link will not be reachable anymore. It makes this method not desirable because the AS cannot exploit the resilience of multihoming.

A variant of the selective advertisements is the advertisement of more *specific prefixes*. This technique relies on the fact that an IP router will always select in its forwarding table the most specific route for each packet (*i.e.* the matching route with the longest prefix). For example, if a forwarding table contains both a route toward 16.0.0.0/8 and a route toward 16.1.2.0/24, then a packet whose destination is 16.1.2.200 would be forwarded along the second route. Therefore one AS can control its incoming traffic by advertising a large aggregate on all links for fault-tolerance reason and advertising specific prefixes on some links at the same time.

This variant improves the resilience of the network, however, two methods share the same drawback that the AS will advertise a number of prefixes larger than required. All these prefixes will be propagated throughout the global Internet and will increase the size of BGP routing tables of potentially all ASes in the Internet. [3] reports that more specific routes constitute more than half of the entries in a BGP table. Faced with this increase of routing table size, several large ISPs have started to implement filters to reject BGP advertisements corresponding to more specific prefixes. It implies that the more specific prefixes will not be announced by those large ISPs and thus the technique will become much less effective.

The length of the AS Path is utilized as the third criteria in the BGP decision process, therefore a possible way to influence the selection of routes by a distant AS is to artificially increase the length of the AS Path attribute by including multiple of its own AS number. This method allows an AS to indicate a ranking among the various route advertisements that it sends. It is referred to as *AS Path Prepending* (ASPP) approach, which will be described in more detail in Chapter 3.

The ASPP approach does not introduce longer prefixes, and at the same time takes the advantage of resilience protection from multihomed connections. However, the ASPP approach is often performed in a *trial-and-error* basis. Many operators believe that it is prone to surprise changes and it is difficult to predict the impact of a prepending behavior given the limited knowledge of the Internet topology and routing policies used by distant ASes.

The last method is to rely on the multi-exit-discriminator (MED) attribute. This optional attribute can only be used by an AS multi-connected to another AS to influence the link that should be used by other ASes to send packets toward a specific destination. It should however be noted that this method can only be utilized to select one link from multiple links to the same AS. Moreover, the utilization of the MED attribute is usually subject to a negotiation between the two peering ASes and some ASes do not accept to take the MED attribute into account in their decision process.

In addition to the above techniques, several ISPs have been using the community attribute to give their customers a finer control on the redistribution of their routes. The community attribute is an optional attribute which can contain several 32 bits wide community values. When the community attribute is used for traffic engineering purposes, predefined community values can be attached to routes in order to request remote routers to perform some actions, such as not announcing the route to a specified set of peers, or prepending the AS Path when announcing the route to a specified set of peers *etc.*.

The communities allow an AS to flexibly influence the redistribution of its routes toward non-directly connected ISPs. Unfortunately, not all ISPs support all communities, and this technique relies on an *ad hoc* definition of community values and on manual configurations of BGP filters. It makes the technique difficult to use and subject to errors. An extension to BGP community attribute called *redistribution community* is proposed within IETF to solve these problems [7]. However, even this extension still suffers from some important drawbacks. For example, this technique requires changes to the attributes of BGP advertisements and any small change to an attribute will force the route advertisement to be redistributed to potentially the entire Internet.

Remarks: Using BGP policies to shift traffic requires extreme care because the influence of policy changes on traffic flows is difficult to predict. In [32], the authors suggest some guidelines on how to perform traffic engineering properly. For example, operators should make policy changes based on large groups of prefixes (*e.g.*, groups of prefixes that have a common origin AS, or other common attributes), limit policy sensitivity to AS path changes by assigning policies based on AS path regular expression matches and assigning local preference based on ranges of AS path lengths, rather than using AS path length as an absolute metric.

2.3 Routing Instability

Routing instability refers to the following scenario: a node (or an AS) makes a decision and advertises a new route to its neighbors, which causes neighbors to change their decisions; then, these nodes (or ASes) withdraw their previous routes and advertise new ones, which in turn causes the AS to withdraw its previous decision and advertise a new decision. This process repeats and continually introduces advertise and withdraw routes to a given prefix, without ever converging to a set of stable routes.

Routing instability is an important problem currently facing the Internet engineering community. It can spread from router to router and propagate throughout the network. The routing instabilities would result in a large number of routing updates, significantly contribute to poor end-to-end network performance and degrade the overall efficiency of the Internet infrastructure. The primary effects include increased packet loss, delays in the time for network convergence, and additional resource overhead (memory and CPU *etc.*). In this section, we will introduce the routing instability problem induced by policy disputes in BGP routing system, and possible solutions for this problem.

2.3.1 Policy Induced Routing Instability

As required of any interdomain routing protocol, BGP allows each autonomous system to independently define its routing policies with little or no global coordination, and it allows local preference attribute to override distance metrics in favor of policy concerns. As a result, the relative ranking of routes in BGP is not, in general, based on path lengths, or any other universally agreed upon cost function, and each autonomous system can reject paths arbitrarily (even shortest paths) based on policy considerations. In the absence of policy coordination, it is possible for different autonomous systems to establish conflicting routing policies that are unsafe in the sense that they can cause routing instability.

There is a famous example called *bad gadget* to illustrate the routing instability induced by policy disputes. In Figure 2.2, suppose there is one single destination prefix d originated by AS_0. The routing policies of AS_1, AS_2 and AS_3 are as follows: Each AS prefers the counter-clockwise route of length 2 to all other routes to d. For example, for AS_3, the route (AS_3, AS_2, AS_0) is assigned the highest local preference, and is preferred to other routes to the prefix d like (AS_3, AS_0) and (AS_3, AS_1, AS_0).

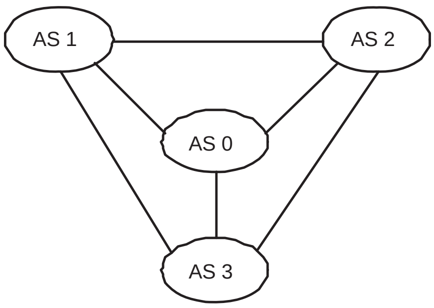

Figure 2.2: The BGP System BAD GADGET

There is no stable routing solution in this network. The ASes in this network set up mutually dependent policies, that lead to non-terminating routing information exchange in BGP. This system is first presented by Kannan Varadhan *et al.* in [21] and then appear in many papers.

In order to solve this instability problem, in [50] the authors present sufficient conditions on routing policies that guarantee BGP safety based on a concept of *dispute cycle*, which is a certain circular set of relationships between routing policies at various autonomous systems. This analysis is based on an abstract routing formalism called Simple Path Vector Protocol (SPVP).

In [22] authors study persistent route oscillation whenever ASes do path attribute based, independent route selection (PAIRS).

The authors derive necessary and sufficient conditions for the existence of route oscillations based on the definition of *return graph* (a representation of domain preference functions). However, this analysis is only for a very restricted class of interdomain topologies.

To study the instability problem formally, Griffin *et al.* propose an abstract model called *stable path problem* to capture the underlying semantics of any path vector protocol such as BGP [51]. In this abstraction, the network consists of a set of processes interconnected by channels. Each process informs its neighboring processes of its current path to the destination. From the paths received from its neighbors, each process chooses the best path according to its local routing policy. Since routing policies are chosen locally, conflicts may occur between processes, resulting in unstable behavior.

The route instability problem with using MED attribute has been studied in [52]. The authors present the first rigorous analysis of the MED instability problem using the stable path problem formalism. They prove that the scope of any MED Induced Routing Anomalies (MIRA) is always contained within a single autonomous system if the BGP configurations between ASes follow a common model based on provider-to-customer and peer-to-peer relationship.

2.3.2 Solutions to Routing Instability

Avoiding the routing instability problem is crucial for the performance of the Internet routing infrastructure. Broadly speaking, there are three complementary approaches to addressing this problem: 1) *static analysis of routing policies*; 2) *dynamic detection*; and 3) *operational guidelines*.

A static solution would rely on programs to analyze routing policies to verify that they do not contain policy conflicts that could lead to protocol divergence [35]. This approach may also involve global coordination effort such as the use of the Internet Routing Registry, which is a repository of routing policies specified in a standard language. However, there are two practical challenges facing this approach. First, autonomous systems currently do not widely share their routing policies, or only publish incomplete specifications. Second, even if there were complete knowledge of routing policies, checking for various global convergence conditions is either NP-complete or NP-hard [49]. Therefore, a static approach would most likely require the development of new heuristic algorithms for detecting this class of policy conflict.

A dynamic solution to the routing instability problem is to suppress or completely prevent at "run time" those BGP oscillations. *Route flap dampening* is an example of this kind of solution. However route flap dampening only makes these oscillations run in "slow motion" and cannot eliminate them. Furthermore, route flap dampening events do not provide network administrators with enough information to identify the source of the route flapping. Route flapping caused by policy conflicts will look the same as route flapping caused by unstable routers or defective network interfaces. It seems that any dynamic solution would require an extension to the BGP protocol to carry additional information that would allow policy disputes to be detected and identified at run time. For example, the *path history* attribute proposed in [48].

Another approach, proposed by Gao and Rexford in [27], is to construct a set of reasonable policy guidelines that guarantee a safe BGP system without requiring coordination between ASes. The approach exploits Internet's hierarchical structure and commercial relationships between ASs to impose a partial order on the set of routes to each destination. Under their guidelines, routing via a peer or a provider is never preferable to routing via a customer link; furthermore, routes via backup links have the lowest preference. An AS is free to apply any local policy to the routes learned from neighbors within each preference class. Hence the goal to avoid conflict is achieved by restricting the set of policies that each AS can apply, so that the desired property of locally choosing a routing policy is lost. Furthermore, it is crucial to make these guidelines explicit since BGP itself does not constrain routing policies to ensure convergence. The authors propose a simple routing registry that stores only the relationship between each AS pair, rather than the entire set of routing policies. These relationships can be explicitly registered by the ASes or inferred from BGP routing tables available throughout the Internet.

2.4 Overlay Networks

Over the past several years, we have seen the emergence of numerous types of "overlay" networks in the Internet to enhance or modify the basic functioning of traffic handling within the Internet. There are many diverse examples of such overlay networks, including the content distribution networks (CDNs) implemented by companies like Akamai, the peer-to-peer file sharing networks associated with applications such as BitTorrent, the voice-over-IP services offered via Skype, and various testbed networks such as PlanetLab.

Overlays emerge for a variety of reasons. First, overlays may exist to support the special requirements of a particular class of application or user community. Second, overlays may arise because of conflicts in stakeholder interests, reflecting a tussle between and among customers, service providers and policy-makers. Third, overlays may play a role in the dynamic evolution of Internet technology. Overlay networks can provide a way to first experiment with new routing and architecture designs (*e.g.*, PlanetLab) and then as a way to incrementally deploy new solutions. These forces are fundamental and enduring, so we expect that overlays will remain an important and growing feature of the Internet.

2.4.1 CDN and Routing Overlay

The two famous Content Distribution Networks (CDNs) are the commercial CDN offered by Akamai and the cooperative CDN based on popular peer-to-peer applications such as BitTorrent. These two kinds of CDNs represent a large share of overlay traffic on the Internet today.

One CDN should consist a lot of caches/peers distributed across the Internet and these caches/peers contain copies of content and services retrieved either on-request or proactively from publishers and providers. When an application requests content or services hosted by a CDN, the CDN services the request from one or more of its distributed servers (or peers) throughout the Internet. The heart of one CDN is the selection algorithm to determine which servers (or peers) should handle the request. Based on various selection algorithms, CDNs optimize different criteria including technical measures such as response time and server loads and economic measures such as bandwidth costs. The techniques that can be used by CDNs include DNS and URL rewriting and http redirection.

Another example of overlay networks is routing overlay. The purpose of routing overlays is to control the path of data through the network at application layer. The routing overlay is unique among classes of overlays because they exist to change the way an existing function is performed, while most other overlays provide new functionality. Routing overlays that seek to improve basic Internet route selection process (*e.g.*, BGP) at application layer may be in conflict with policy-based routing implemented by ISPs in response to other non-delay-related considerations (*e.g.*, long-term interconnection agreements or regulatory-jurisdiction issues). For example, an overlay that tries to select the "best" route based on global information about link delays may violate business agreements between ISPs that are seeking to manage traffic to minimize intercarrier payments.

In routing overlays, the endpoints of the information exchange are unchanged from what they would have been in the absence of the overlay, but the route through the network that packets traverse between these endpoints may be different - the resultant path depends on the application layer routing by the overlay as well as the network layer routing by the basic Internet. In CDNs, the IP layer is still responsible for delivering the packet to the appropriate destination but the decision about the source of packets is made at the application layer by the redirector, and not the original requestor. It can also be viewed as a kind of *application layer routing.* With application layer routing, the control of the route selection is, at least to some extent, wrested away from network operators and shifted to end users. This loss of control over a basic function of network operation has strong implications for the interests of ISPs.

2.4.2 Impact of Overlays

The current architecture historically has been based on the "end-to-end" principle which relied on a relatively clear demarcation between applications and network services, and edge and core responsibilities. However, overlays exist in the blurry boundary between what we think of as "the Internet" (a globally interconnected network of IP networks) and the applications that exist on top of the Internet. Overlays also blur the boundary between network edges (customer end-nodes) and the network core (the services that support the Internet). As such, overlays have important technological and policy implications for the evolution of next generation Internet architecture.

The overlays change the patterns of inter-ISP traffic flows by application layer routing, so they have a clear technical impact on the Internet. Since inter-ISP traffic flows also determine money flows between providers, overlays also influence commercial relationships on the Internet, which in turn, gives rise to policy implications.

Let us illustrate these influences in more detail. For example, CDNs result in a larger proportion of traffic being served locally, incentives to and the need for investments in long-haul end-to-end capacity will be reduced. It reveals the technical influence of overlays on capacity planning and infrastructure investment. Economically, because of the efficient caching at access ISPs, regional and backbone ISPs will carry less traffic and face lower revenues. Peering backbone ISPs may also face lowered aggregate traffic volume across their peering points. Particularly, whether traffic reductions at the peering point are symmetric - thereby maintaining the peering contract condition on which "free peering" is based - is conditional on the presence and selection algorithms of CDNs on both sides of the peering point. Thus, one CDN (assuming it carries a significant fraction of Internet traffic) can potentially unravel the peering relationship or help correct existing imbalances. Overlays also pose challenges for policy makers in multiple dimensions. For example, if CDNs provide superior access selectively to some content, would this give rise to a two-tiered Internet: one with high quality for commercial content and one with lower quality for non-commercial content? If so, would this raise concerns about equal, non-discriminatory access? Suppose an ISP sought to vertically integrate with a major CDN provider like Akamai, would that raise antitrust concerns?

We see that many interesting questions raised by overlays are not only technical; instead they are also related to the changing relationship between ISPs and their customers. As we stated at the end of Section 2.1, before overlays appear the route is determined entirely by ISPs, without input or control from end users or applications. However, end users may have different concerns from ISPs. As long as ISPs retain complete control over routing decisions within the network, there is little call for technical routing mechanisms to resolve the “tussle” between the choices of ISPs and those of end users. Overlays change this equation by giving users an input into the routing decision. However, currently they do not provide a coordinated way to resolve conflicting objectives between various parties. Instead, they simply allow end users to override ISPs in certain situations. It is this lack of coordination that leads to many of the negative effects of overlays.

Overlays gain widespread significance as a method for end users and third parties to affect routing decisions and they become vehicles for contention over the routing decision between ISPs and end users. The use of overlays may create substantial effects for ISPs, and potentially for the overall stability of the Internet [53] [60]. At minimum, additional research and development are required before overlays can safely fulfill this role, especially on technical and economic mechanisms that better coordinate the multi-party routing decision.

□ **End of chapter.**

Chapter 3

Interdomain Traffic Engineering

Summary

In this chapter, we restrict our attention to the *AS Path Prepending* approach to study various issues in interdomain traffic engineering. In order to improve the current situation that ISPs often practise this approach in a trial-and-error basis, we propose a greedy algorithm to help ISPs perform this approach systematically and efficiently. Then we demonstrate some fundamental issues of decentralized selfish traffic engineering in the Internet, such as *routing instability* and *global network performance degradation*, based on an abstract model where ISPs perform traffic engineering for their individual load balance. We also present a real-world pathologic case of prepending instability from our measurement study of the dynamic prepending behavior of ISPs in real Internet. Some simple guidelines are given for ISPs to avoid such routing instability.

The focus of this chapter is on the *inbound interdomain traffic engineering* problem, which is known to be more difficult than the outbound traffic engineering problem because an AS generally cannot control the routing path for inbound traffic directly. Moreover, we restrict our attention to the *AS Path Prepending* (ASPP) approach.

The application of the ASPP approach has been introduced in Section 2.2. Here we further illustrate it using the following example. Consider the traffic from AS_1 to AS_5 in Figure 2.1. In this network, AS_1 receives two routes for prefixes in AS_5: $(AS_4, AS5)$ and (AS_5). These two routes have the same local preference because both of them are announced by AS_1's customer neighbors (AS_5 and AS_4), then the router in AS_1 selects the second route as the preferred route for prefixes in AS_5 since it has a shorter AS path. If AS_5 wishes that traffic from AS_1 goes through the link (AS_4, AS_5), it can use ASPP and announce AS path (AS_5, AS_5, AS_5) to AS_1. Now AS_1 receives two routes with AS path (AS_4, AS_5) and (AS_5, AS_5, AS_5). Therefore, the router in AS_1 would change its decision and choose the first route as the best route.

Particularly, an AS can ask other ASes to do prependings for it through community attribute. After ASes make agreements on the meaning of some special identifier values, community attribute can be used by an AS to affect routing policies of other ASes. For example, AS_6 sends an announcement to AS_4 with a pre-defined community value $400 : 002$. After AS_4 receives this route and extracts this community value, it knows AS_6 wants it to add its AS number three times (prepend the route by 2) when it advertises this route to AS_0. Therefore, AS_4's routing policy is affected by AS_6, and AS_0's best route selection could be affected by AS_4's prepending action.

3.1 Popularity and Justifications

According to our measurement study, 33% of multihomed stub ASes use ASPP for inbound traffic control, and the share of such transit ASes from the total number of transit ASes is around 40% in 2004. We also find that the share of prepended routes has been increased to more than 12% in 2004. Moreover, these percentages exhibited an increasing trend in the last few years. These results indicate this approach has been widely employed in the Internet today [18].[1]

Since ASes have different traffic distribution targets due to their own technical and business requirements, it is difficult to discern ASes' concrete goals of their ASPP behaviors. Based on our discussions with some network operators, we summarize some justifications of performing ASPP as follows:

- *Load Balancing*:

  The prepending is performed because the AS wants to balance its inbound load to meet its capacity requirement. For example, assume AS_a has a Gigabit Ethernet (GigE) link to one provider, and an OC3 to another. Then AS_a is likely going to prepend the OC3 link. Likewise, if inbound traffic is pushing the AS over its minimum commitment bandwidth on one provider, but is well under its minimum on another, the AS may prepend to help balance the traffic levels.

- *Cost Minimization*:

[1] We need to point out that the result only shows a conservative view of prepended routes in the Internet. As we know, a BGP router only advertises to its neighbors the routes that are selected as its best routes. So routes with prepending, especially with higher number of prependings, were most likely not selected by transit ASes and therefore filtered out. So the fact that we observe so many prepended routes in Route View's routing tables also implies ASPP does not always produce intended results for those ASes that are included as prepended ASes in the AS paths. Otherwise, such paths would not be present as the best paths in the routing tables.

In order to minimize the transit cost, a multihomed AS may want to achieve a particular traffic distribution. For example, AS_c has two providers, AS_a and AS_b. The total inbound bandwidth of AS_c is 15MBps. The cost for the traffic going through AS_a is \$10/MBps, while the cost for the traffic going through AS_b is \$8/MBps when the bandwidth is below 10MBps, \$12 when the bandwidth is over 10MBps. Thus AS_a wants to tune the inbound traffic to achieve this traffic distribution: 10MBps on the link to AS_b and 5MBps on the link to AS_a. It can be implemented through the ASPP approach.

- *Performance Optimization*:

  In general, the length of AS path is not a good metric to measure the performance of a path, *e.g.*, although a route via AS_a has the shortest AS path, the performance of this path may not be the best [5]. Then one AS might prepend this kind of paths to achieve a better performance.

- *Creating Backup Route*:

  Some links only serve as backup paths. One AS may want to prepend a link to make this path a backup choice for failover purpose. In this case, the AS would increase the prepending length on the link until no traffic can be shifted.

The popular employment of the ASPP approach should have a significant impact on the current Internet routing structure. However, it is surprising that there still lacks a systematic study of this approach and the basic understanding of its effectiveness.

3.2 Terminologies and Notations

We represent a route simply by its AS path since only this attribute is related to our interest. Under BGP convention, an

AS paths is represented by a sequence of AS numbers. An AS path $(a_0, a_1, \ldots, a_n)$ means for $i \in [1, n]$, AS_{a_i} announces the route $(a_i, \ldots, a_n)$ to $AS_{a_{i-1}}$, then $AS_{a_{i-1}}$ learns that AS_{a_i} agrees to transit traffic destined to AS_{a_n} for $AS_{a_{i-1}}$ through the AS sequence $(a_i, \ldots, a_n)$. A prepended AS path is an AS path that has some *duplicated* AS numbers that appear consecutively.

For convenience, an AS path can also be represented as a sequence of links it traverses $(e_1, e_2, ..., e_n)$. For a prepended AS path, the corresponding links are prepended. For example, $((0, 4), (4, 6))$ is equivalent to $(0, 4, 6)$; and $((0, 4), (4, 6), (4, 6))$ is equivalent to $(0, 4, 6, 6)$. In this case, we say "link $(4, 6)$ is prepended once".

Given a source and destination pair (s, d) $(s, d \in V)$, let $R_{s,d}$ be the set of all routes from s to d allowed by business relationships $B(\cdot)$; and $R^*_{s,d}$ to denote the set of all shortest routes. To illustrate, consider again the graph in Figure 2.1 with the source node being AS_0 and the destination node being AS_6. We can see that $R_{0,6} = \{(0, 2, 3, 6), (0, 4, 3, 6), (0, 4, 6)\}$. The shortest path set in this example is $R^*_{0,6} = \{(0, 4, 6)\}$.

let $E(v)$ be those links connecting AS_v to its providers. ASPP is performed by a multihomed AS_v on its links $E(v)$. There are three types of prepending policies that a multihomed AS can employ:

- *link-based prepending*: An AS_v is said to employ a link-based prepending policy if the prepending length (including 0 for non-prepended routes) is the same for *all* routes announced to a specific link. However, the prepending length may be different across the links, *e.g.*, one link without prepending and the other with a prepending length of 3.

- *destination-based prepending*: An AS_v repeats a link $e \in E(v)$ for all routes that traverse link e and destine for AS_d. It means that AS_v performs prepending for all prefixes of one AS in the same way. An AS can perform prepending on its own routes, *i.e.*, $AS_v = AS_d$, or an AS_v can perform prepending on routes generated by its customer or descendant customer AS_d.

- *prefix-based prepending*: It is the "*finest-granularity*" prepending policy, where the prepending lengths of routes that traverse a link $e \in E(v)$ and destine for a same AS are different.

For stub ASes, the only incoming traffic over its provider links are destined for the stub AS, hence we make the following observation according to the above definitions:

Observation 1 *For a stub AS, destination-based prepending is the same as link-based prepending.*

In [18] the measurement result shows a downward trend on the percentages of the ASes that use link-based prepending. Similarly, the percentages of the links that involved link-based prepending policy also show a downward trend. It is apparent that prepending policies are becoming more complex than a "lazy prepending" approach.

However, currently link-based prepending is still very popular since it is easier to configure and implement. [18] shows that more than 70% of multihomed stub ASes employ link-based prepending policy, and the share of links involved link-based prepending is around 80%. So in the rest of this chapter, we focus on link-based prepending and drop the word "link-based".

Assuming that AS_v repeats the link $e \in E(v)$ n times for all routes that traverse e, we represent this link-based prepending by $\mathcal{P} = e^n/*$, or simply $\mathcal{P} = e^n$. The set of routes after this prepending action is represented by $R_{s,d}(e^n)$. If the prepending is only for routes destined to AS_d (destination-based prepending) or prefix p (prefix-based prepending), we represent the action by $\mathcal{P} = e^n/d$ or $\mathcal{P} = e^n/p$. The set of routes after this prepending action is represented by $R_{s,d}(e^n/d)$ or $R_{s,d}(e^n/p)$.

Consider in Figure 2.1 that AS_6 tries to reduce the amount of traffic on its incoming link from AS_4 by prepending *twice* on this incoming link. In this case, $\mathcal{P} = \{(4,6)^2\}$. The set of routes after this prepending operation is represented by $R_{0,6}(\mathcal{P}) = \{(0,2,3,6),(0,4,3,6),(0,4,6,6,6)\}$. The set of shortest paths from AS_0 to AS_6 is now $R^*_{0,6}(\mathcal{P}) = \{(0,2,3,6),(0,4,3,6)\}$.

3.3 Greedy ASPP Search Algorithm

Although the ASPP mechanism is widely used in ASes' traffic engineering for various goals, there is little prescription for a systematic way to implement it. How does an AS find the prepending action that optimizes its local objective? Obviously it is infeasible to try all combination of prepending lengths on all provider links up to some limit for each link. Here, we propose a *Greedy ASPP Search Algorithm* for ISPs to practise ASPP systematically and efficiently.

3.3.1 A Greedy ASPP Search Algorithm

In today's Internet, ASPP is purely a heuristic method. ISPs do it on a trial-and-error basis, which may take some time to converge to a desirable ASPP configuration and in the meantime make real customer traffic try out different routes. In [38], the authors propose a systematic procedure to predict the changes in traffic distribution for a given new ASPP configuration. Assume AS_v have m provider links, and the current (incoming) traffic distribution be $t(p) = (t_1, t_2, \ldots, t_m)$ where p represents the current ASPP configuration and t_i denotes the traffic intensity on the *ith* provider link. The procedure in [38] would then predict the new traffic distribution $t(p')$, where p' is the new

ASPP configuration. In this thesis, we refer to this procedure as *ASPP Impact Estimator*. This estimator works as follows:

1. use passive traffic monitoring (such as netflow analysis) to identify a few top (heaviest) traffic flows;
2. announce BGP routes for an unused IP prefix a in AS_v with the new ASPP configuration p';
3. after new BGP announcements take effect, ping typical source addresses (representing the top senders identified above) from a, and watch for any change in the routes for these top flows;
4. based on the change in the routes for the top flows, estimate the change in the volume of traffic, $t(p')$, by assuming the ratio of route change for other flows is the same as that for the top flows.

In this algorithm, a is sometimes referred to as a BGP beacon [61]. Since the beacon is deployed in AS_v, this procedure only works for estimating shifts in traffic destined for the local AS AS_v. Therefore, it is most suitable for a stub AS. According to our measurement, 50% of the ASes today are multihomed stub ASes which are the prime candidates of using ASPP to control the inbound traffic coming into their links.

If an AS has a well-defined traffic engineering goal (*viz* $t^* = (t_1^*, t_2^*, \ldots, t_m^*)$), it is theoretically possible to search for the best ASPP configuration p^* that best meets the target traffic distribution. Based on the prediction of the ASPP Impact Estimator, this search procedure can be implemented without affecting the traffic in the real Internet. We propose that ISPs can deploy the following search algorithm, which is to be referred to as the *Greedy ASPP Search Algorithm*.

To describe the algorithm more formally, let t^* denote the desired (optimal) traffic distribution. Let $f(t(p))$ be a measure of the goodness of a given traffic pattern t, resulting from p. By definition, $f(t(p)) \leq f(t^*)$. Given p in which link e is prepended, let $p-e$ denote the prepending configuration with one prepending on e removed; similarly let $p+e$ denote p with one additional prepending on link e.

Greedy ASPP Search Algorithm:
ASes execute this algorithm to search for the best prepending configuration

```
while ( TRUE ) {

  compute f(t(p));

  let e be the link with most room to add traffic according
  to the desired distribution;
  if (the prepending length on e > 0) {
    p' = p - e;
    if (f(t(p')) > f(t(p))) {
      p = p - e;
      continue;
    }
  }

  let e be the link with most room to reduce traffic
  according to the desired distribution;
  p'' = p + e;
  if (f(t(p'')) > f(t(p))) {
    p = p + e;
    continue;
  }

  break;
```

18. }

The algorithm first tries to reduce the prepending length on the lightest-loaded link, then tries to increase the prepending length on the heaviest-load link. Similar as other greedy search algorithms, the basic idea of this algorithm is to search in the most likely helpful direction in each step, until the desired traffic distribution is reached or there is no helpful prepending action.

3.3.2 Example: Single AS Load Balancing

The application of the Greedy ASPP Search Algorithm is illustrated using the network in Figure 3.1. For our purpose here, we assume that AS_1 (and only AS_1) is doing load balancing via the greedy algorithm. First, we need to define *local load balancing index*, $\mathcal{I}_{lb}(v)$, to model how AS_v evaluates its traffic distribution.

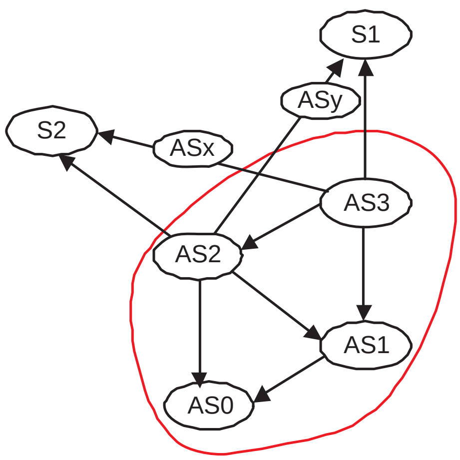

Figure 3.1: Example Network for Studying Prepending Algorithm

Consider a single multihomed AS_v, $\mathcal{I}_{lb}(v)$ measures the degree of load balancing of incoming traffic on provider links $E(v)$. If an AS has two provider links, the answer is quite straightforward since it is intuitive that a traffic ratio of 2:3 is better than a traffic ratio of 1:4. If an AS has more than two provider links, then one has to define the degree of load balancing carefully. Let say t_e is the traffic intensity on a link $e \in E(v)$, and c_e is the bandwidth of the link e. We define $I_{lb}(v)$ as follows:

$$f(t(p)) \;=\; \mathcal{I}_{lb}(v) = \frac{\left(\sum_{e\in E(v)} t_e/c_e\right)^2}{|E(v)| \sum_{e\in E(v)} (t_e/c_e)^2}, \tag{3.1}$$

This index was first proposed for measuring the fairness of bandwidth allocation [33], but it also serves the purpose to measure the degree of load balance. Note that $\mathcal{I}_{lb}(v) \in (0, 1]$. When $\mathcal{I}_{lb}(v)$ is close to 0, it implies the traffic loading on $E(v)$ is very skewed. When $\mathcal{I}_{lb}(v)$ is close to 1, it implies the loading of $E(v)$ is closely balanced. Without loss of generality, we assume all provider links have the same bandwidth unless we state otherwise.

For all examples in this chapter, we do not model the effect of the ASPP Impact Estimator. Instead, we calculate the traffic distribution directly using the following Equation 3.2, which means we assume that the ASPP Impact Estimator can give the exact resultant traffic distribution after prepending.

Let T denote the relative traffic demand matrix, where T_{ij} represents the *relative* end-to-end traffic demand between the source node $i\in V$ and the destination node $j\in V$. The traffic on a link $e \in E$ is simply the sum of traffic, using the shortest path route allowed by the business relationships, that traverses link e. That is:

$$t_e \;=\; \sum_{s\in V}\sum_{d\in V}\sum_{r\in R^*_{s,d}(\mathcal{P})} \frac{1}{k} T_{sd} I_{e\in r}. \tag{3.2}$$

where I is an indicating function with value of either 0 or 1, T_{sd} is the demand from the source AS_s to the destination AS_d, and $k = |R^*_{s,d}(\mathcal{P})|$, which is the number of shortest paths in the set.

When the shortest route set $R^*_{s,d}$ consists of a set of (more than one) paths, we assume traffic from s to d is evenly divided on these paths. This assumption tends to balance traffic automatically, because we are interested in studying how load balancing works even under this more *favorable* assumption.

In the network in Figure 3.1, there are four nodes, representing four ASes: AS_0, AS_1, AS_2 and AS_3. There are other nodes solely for the purpose of generating and forwarding traffic. In particular, $S1$ and $S2$ are nodes which will generate traffic to AS_0 and AS_1. Note that in the next section, we will use the same network to illustrate the convergence issue and performance issue when multiple ASes are doing prepending. Since there is no other traffic in this network, the traffic matrix T can be simply represented by a two by two matrix, specifying the relative traffic intensity from $S1$ to AS_0 and AS_1, and $S2$ to AS_0 and AS_1 respectively. Consider the following traffic matrix:

$$T = \begin{pmatrix} 140 & 10 \\ 0 & 10 \end{pmatrix}.$$

The operation of the Greedy ASPP Search Algorithm by AS_1 in this network is summarized in Table 3.1.

$\mathcal{P}$	$t_{(2,1)}$	$t_{(3,1)}$	$\mathcal{I}_{lb}(1)$
$\varnothing$	10	56	0.67
(3,1)	16	3	0.68
(2,1)	10	10	1.00

Table 3.1: AS_1 Executing the Greedy ASPP Search Algorithm for Load Balance.

When there is no prepending, the amount of traffic on link $(2,1)$ is 10 and on link $(3,1)$ is 56. Note that *some* of the traffic on link $(3,1)$ is destined for AS_0. When AS_1 executes the Greedy ASPP Search Algorithm, it first chooses the link $(3,1)$ to prepend and the load balance index increases to 0.68. It then executes the algorithm again and prepend the link $(2, 1)$. After this prepending operation, no further prepending is necessary since there is $\mathcal{I}_{lb}(1) = 1$ which means these two links are perfectly balanced.

3.3.3 Performance Analysis of the Algorithm

Now let us conduct an analysis on the performance of the Greedy ASPP Search Algorithm:

Observation 2 *The greedy ASPP search algorithm stops after a finite number of iterations for any single AS.*

Consider AS_v with m providers, and use a vector $(l_1, l_2, \ldots l_i \ldots l_m)$ to represent one of its prepending configuration, where l_i is the prepending length on its ith provider link in this prepending configuration.

Let n be the diameter of the network which is the length of the longest AS path among all possible paths within this network. Obviously, prepending a link with a length of n should be enough to shift all traffic whose routing can be affected by ASPP on this link to other links, therefore prepending a link with a length of more than n should have the same effect with a prepending with a length of n. In other words , the maximum useful prepending length in this network is n, or we can say that n is the upper bound of the length of a useful prepending.

So we can assume $0 \le l_i \le n (i = 1 \ldots m)$. For AS_v, there are $(n + 1)^m$ possible prepending configurations. Let us sort all these prepending configuration as $(p_1, p_2, \ldots, p_i, \ldots, p_{(n+1)^m})$, where $f(t(p_i)) \le f(t(p_j))$ when $i < j$.

During the execution, the greedy algorithm would generate a series of prependings configurations $P = (p_{a_1}, p_{a_2}, \ldots, p_{a_i} \ldots p_{a_k})$ where p_{a_i} is the prepending configuration after ith iteration. $f(t(p_{a_i})) < f(t(p_{a_j}))$ must hold for any $i < j$ because the algorithm should improve the traffic distribution in each iteration.

So we have $a_1 < a_2 < \ldots < a_k \leq (n+1)^m$. We can see that k should be a finite integer which means the greedy ASPP search algorithm stops after a finite number of iterations for any single AS.

Observation 3 *The result of the greedy ASPP search algorithm is not guaranteed to be optimal, but the performance is good and the algorithm converges fast.*

The result of the greedy algorithm is not guaranteed to be optimal because the AS only tries prepending changes on two links in each iteration, while other links are ignored. However, our simulation shows that the performance of this algorithm is really acceptable and it converges fast.

We do the simulation as follows. We focus on one stub AS, say AS_v. The goal of AS_v is to balance its incoming traffic using ASPP approach. We assume all links in the network have the same bandwidth in order to simplify the simulation. AS_v measures the degree of load balance on its provider links by Equation 3.1.

In each simulation run, we generate a set of top 100 flows destined to this AS. The intensity of these flows obeys the "power law" distribution. Particularly, we generate the flow intensity according to the measurement data shown in Figure 3 of [46] and Figure 12 of [32]. Each flow can reach AS_v via multiple paths. The distribution of their path lengths is based on the measurement data of Cernet IPv4 [1]. We apply the greedy algorithm to search for the best prepending configuration. We then compute the optimal configuration by exhaustive search and compare this optimal configuration with the result from the greedy algorithm. After 3000 runs, we summarize the results in terms of cumulative distribution functions of different outcomes, as shown in Figure 3.2 and Figure 3.3.

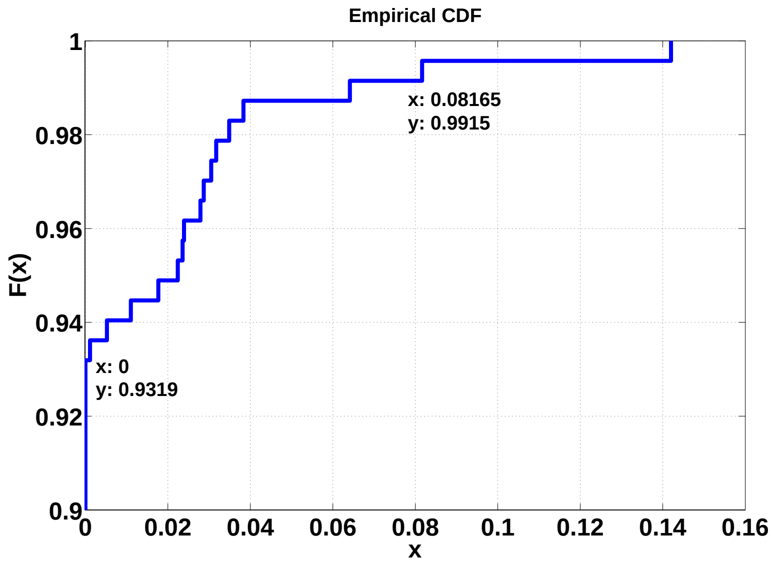

Figure 3.2: Comparison of the Result from the Greedy Algorithm with the Optimal Result

In Figure 3.2, we compare the resultant load balance index from the greedy algorithm with the result from the optimal configuration (derived by exhaustive search). Let p_o denote the optimal prepending configuration, and p_g denote the prepending configuration resulting from the greedy algorithm. We plot the cumulative distribution function of $x = \dfrac{f(t(p_o)) - f(t(p_g))}{f(t(p_o))}$ to evaluate the performance of our greedy algorithm.

We can see that the greedy algorithm gives the optimal balance in about 93.19% simulations, and $x > 0.08165$ only in 0.85% simulations. It shows the performance of our algorithm is acceptable.

In Figure 3.3, we plot the cumulative distribution function of the iteration numbers of 3000 simulation runs. We can see that 9.362% simulations stop in three iterations, 98.3% simulations stop in four iterations, and only less than 1.7% simulations take more than 4 iterations to converge. This result shows our algorithm converges fast.

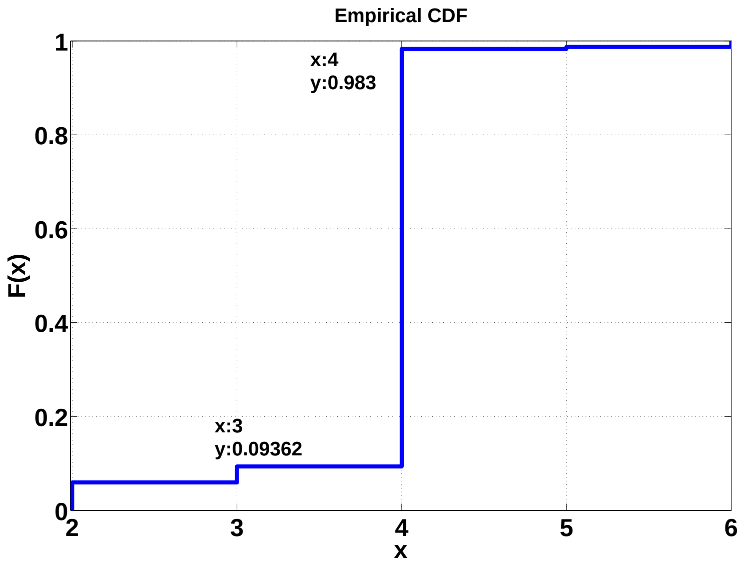

Figure 3.3: The Distribution Function of Iteration Numbers

An exhaustive search algorithm can find the optimal prepending configuration. However, it is infeasible in practice, especially when the ISP has many provider links. There are at least two reasons for this infeasibility. First, there are many prepending configurations so that the AS cannot try all of them. Second, the procedure to evaluate the impact of each ASPP configuration can be very tedious and time consuming.

Currently ISPs always perform ASPP by trial-and-error, and the process may produce a lot of route update messages. Our proposed algorithm is a feasible approach for ISPs to perform ASPP systematically and efficiently.

3.4 ASPP by Multiple ASes

Internet is a large-scale distributed system. The selfish or non-coordinated behavior of nodes in a distributed system often introduces performance or instability issues. Because BGP enables ASes to independently control the traffic across their network based on local objectives and local information with little or no global coordination, ISPs' behavior in the Internet may also cause routing instability or performance degradation problems.

Various kinds of BGP route oscillation problems have been studied in the past. But none of the previous works considered the interaction of ASPP actions of different ASes in the Internet. In this section, we will study the scenario where multiple ASes are performing prepending actions at the same time, focusing on the global effect of ISPs' local actions. First, we make the following observation.

We define a *network cut* as a set of links when removed will partition the network into two halves. Further, if each link in the cut is prepended by the same number of times, we call it a *uniform cut prepending.*

Observation 4 *Let E_{cut} represent a cut of the network graph G, and $\mathcal{P} = E_{cut}^{n}$ be a uniform cut prepending. Then $R^*_{s,d}(\mathcal{P}) = R^*_{s,d}(\varnothing)$ $\forall s, d$. In other words, a uniform cut prepending has no effect on routing in the network.*

At the beginning of this chapter, we list some reasons for ISPs to perform ASPP. Generally speaking, these justifications can be classified into two categories *i.e.* 1) load-independent reasons such as backup; 2) load-dependent reasons such as load balancing. Both categories of prependings can cause routing instability problem. We will discuss them one by one in the following subsections.

3.4.1 ASPP for Load-independent Reasons

Let us consider the network topology in Figure 3.1 again to demonstrate the interaction of ASPP by multiple ASes for load-independent reasons. Assume that AS_2 is more expensive than

AS_1 for AS_0, therefore AS_0 would like to make the link (AS_2, AS_0) a backup link using the ASPP approach. Similarly, the link (AS_3, AS_1) is used as a backup link by AS_1.

In this case, AS_0's prepending policy is to "increase the prepending length on the link (AS_2, AS_0) until no traffic can be shifted to the other link", while AS_1's prepending policy is to "increase the prepending length on the link (AS_3, AS_1) until no traffic can be shifted to the other link". We can find these two local policies can be satisfied at the same time. The solution is that the traffic from other nodes to AS_0 goes through $(\ldots, AS_2, AS_1, AS_0)$. So after AS_0 and AS_1 find the needed prepending configurations and perform their prepending actions, the network becomes stable. Here, the local policies of these ASes in the network are not conflicting.

However, the distributed prepending actions under conflicting policies of multiple ASes may interfere with each other and make the routing unstable since there lacks global coordination. For example, assume that AS_0 in the network would like to avoid its traffic going through AS_2 for competition or performance reasons. Then AS_0 itself would prepend the link (AS_2, AS_0) directly to make this link a backup path. It would also ask its providers (AS_1 here) to prepend the routes for its prefixes when sent to AS_2 using community-based prepending methodology. So AS_0's prepending policy is "do prepending directly and indirectly until no traffic via AS_2 can be shifted", while AS_1's prepending policy is "do prepending until no traffic on the link (AS_3, AS_1) can be shifted". In this case, the prepending policies of these two ISPs are conflicting and there is no stable routing solution. In this small network, it is easy for AS_1 to detect this conflict. However, it is very difficult to detect this kind of conflicts in the large scale Internet. Routing instability caused by conflicting policies is the same as the *bad gadget* in Figure 2.2.

3.4.2 ASPP for Load-dependent Reasons

In this subsection, we consider the scenario where the ASPP approach is employed by ASes for load-dependent reasons. First, we observe that under certain scenarios, one AS's local prepending actions *do not affect* or *interfere with* the traffic distribution of other ASes.

Observation 5 *If only stub ASes are performing prepending on their local provider links, there is no interference in the network.*

This is clearly true because as one stub AS, say AS_d, doing local prepending, only the traffic on links $E(d)$ is involved and no provider links of any other stub AS is part of the routes to AS_d.

However, when a customer AS_a and its provider (or ancestor provider) AS_b are both doing ASPP to optimize their own traffic distribution, there can be interference in the network. Imagine that AS_a finds a prepending action that can make its traffic distribution better. AS_a conducts this prepending for its own purpose, which at the same time affects AS_b's traffic distribution; it then triggers that AS_b performs a new prepending, and this prepending at the same time changes AS_a's traffic distribution, which finally causes that AS_a performs a new prepending. The process may repeat and continually introduce prepending configuration changes, without the routing in the network converging to a set of stable routes. Whether this instability scenario appears in a network depends on the end-to-end traffic demand matrix, network topology and also ISPs' traffic engineering goals.

In this subsection, we will demonstrate the instability issue described above using some examples. We will also evaluate the effect of local prepending actions on the global network performance based on some predefined measures. Particularly, we explore the following two interesting questions:

- If each AS_v conducts local prepending to improve its local load balancing objective, does the global network be able to reach a stable routing?
- If the network reaches a stable routing, is the result optimal or better than before based on global metrics?

To answer these questions, we model rational ISP behavior as to perform ASPP to balance its inbound traffic with a well-defined local objective, *i.e.* these ASes evaluate their traffic distribution using the index defined in Equation 3.1.

Global Performance Metrics

Intuitively, local prependings may have two global effects. First, they may increase the total amount of inter-AS resource consumption in the network since traffic no longer follows shortest AS paths. Second, a more interesting global effect is the degree that traffic is shifted from congested links to under-utilized links on a network-wide basis. In other words, local traffic load balancing may globally lead to a network that appears to have higher capacity, that is, able to support more users or traffic. In order to evaluate the global influence of prepending actions by multiple ASes, we define the following performance metrics.

- **aggregated resource consumption** A: Aggregated resource consumption, A, is simply the sum of traffic intensities on all links in the network. We have

$$A = \sum_{e \in E} t_e, \tag{3.3}$$

where t_e is given in Equation 3.2. Obviously, A measures the total amount of resource consumption. Note that for a given graph G, the value A can be different after a prepending action.

- **global scalability bottleneck** B**:** This measures the maximum bandwidth utilization factor in the network G:

$$B = \max_{e \in E} \frac{t_e}{c_e}. \tag{3.4}$$

- **global load balancing index** $\mathcal{G}_{lb}$**:** Like the local load balancing index, the global load balancing index also measures the closeness of traffic load on different links:

$$\mathcal{G}_{lb} = \frac{\left(\sum_{e \in E} t_e/c_e\right)^2}{|E| \sum_{e \in E} (x_e/b_e)^2}. \tag{3.5}$$

  Again, $\mathcal{G}_{lb} \in (0, 1]$ and $\mathcal{G}_{lb}$ is close to 1 (or 0) when the traffic on all links in the network is closely balanced (or highly skewed).

One of the interesting global effects we want to study is how much local prepending actions for local load balancing are able to relieve congested links so that the network as a whole can accommodate more users or traffic. Both metrics, the *global scalability bottleneck* and *global load balancing index*, can be used towards this end.

The global scalability bottleneck gives the *"hard"* limit of how many logical units of the traffic matrix the network can support. Further scaling up of the traffic matrix will cause the global bottleneck link to first exceed its capacity. The global load balancing index, on the other hand, specifies a *"soft"* objective that is to make sure traffic is as evenly spread as possible so that no link will be a specially bad bottleneck. Maximizing this latter index is more robust against uncertainties in the traffic matrix. Therefore, to improve the effective network capacity for a certain traffic matrix, we can minimize the global scalability bottleneck and maximize the global load balancing index.

Although we do not specifically try to optimize the aggregated resource consumption metric A, it is a useful gauge to check how much efficiency is lost due to load balancing. Note that in general, prepending will increase the value of A since ASes are selecting longer AS paths for routing.

Simulation Tool

Since the problem at hand is quite involved and one cannot carry out controlled experiments in the Internet, we choose to develop a simulation tool to study various scenarios and carry out performance evaluation so as to observe the local and global effects of AS path prepending.

For small topologies, this simulation tool can be used to visualize the network topology, business relationships between ASes, the routes for different destination (or source/destination pairs) and optimal routes, and display various performance metrics as one carry out prepending operations. This is a pleasant alternative to manually examining various examples. Figure 3.4 shows a screen shot of the GUI of this tool.

The simulation tool is an object-oriented toolkit with various objects performing specific functions. Some of the important objects (or modules) are:

- *Topology Generator:* The module generates various topologies based on different rules (*e.g.*, power-law). It also allows a user to input or draw small scale topologies for testing and visualization.

- *Traffic Generator:* Given the topology, this module outputs the traffic matrix T. Currently, user can specify various rules for traffic generation such as uniform, random or specific input traffic matrix for testing.

- *Route Selection:* This module implements the route selection under the business relationships in the network. In

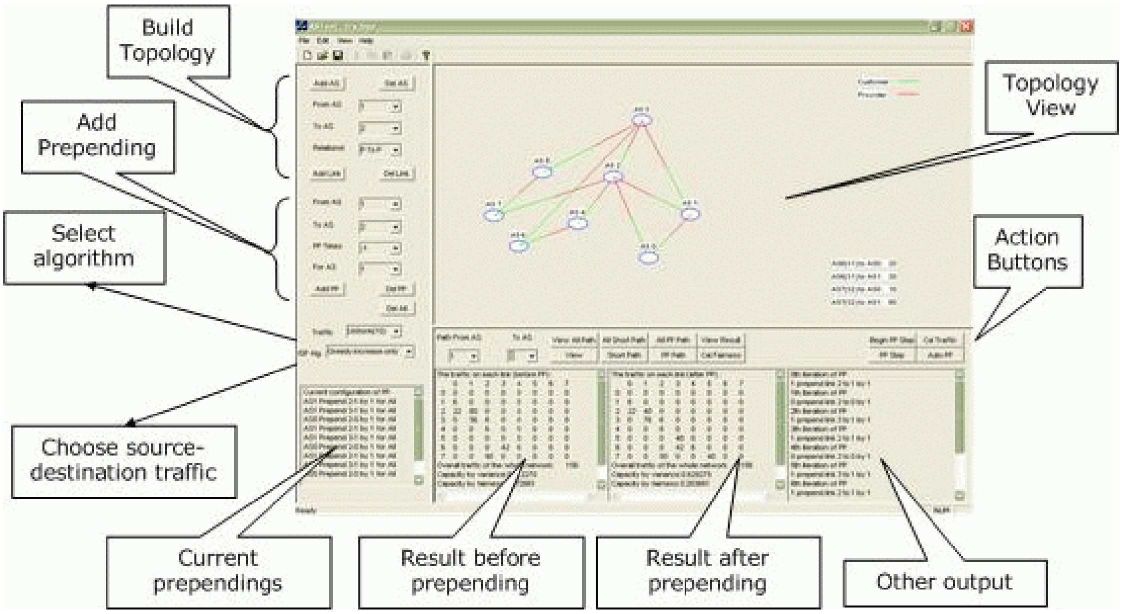

Figure 3.4: Screen Shot of the GUI of Simulation Tool.

particular, popular policies are implemented such as the typical local preference policy and selective export policy we mentioned in the previous section.

- *Prepending Algorithms:* User can select different prepending algorithms. The tool can simulate the process of every AS doing prepending in a synchronous manner and can evaluate different performance measures for different topologies and traffic matrices.

- *Statistic Gathering:* This module is to gather data about various traffic across all links and compute various performance measures mentioned above.

- *Visualization :* This module not only can display the topology under testing but also display various information for debugging. The available information includes:

  – Display all possible paths from a specific source to a specific destination.

– Display the set of shortest path from a specific source to a specific destination, either before or after a prepending operation.
– Display the intensity of traffic on each link in the given topology.
– Display various performance measures (*e.g.*, load balancing index, aggregated resource consumption, global scalability bottleneck ...) before or after a prepending operation.

Now we are going to describe some example scenarios we studied using this simulation tool, as well as some important observations we made. We believe these observations are crucial to understand the stability and performance issues of distributed AS path prepending.

Examples

We use the network in Figure 3.1 here to study the interaction of ASPP actions by multiple ASes. For easy reading, that network topology is presented in Figure 3.5 again. In the following examples, we assume the bandwidth of all links in this network is the same, and $c_e = 100$ ($e \in E$). AS_0 and AS_1 are doing prepending for their inbound load balance on their provider links using the Greedy ASPP Search Algorithm. The degree of load balance is measured by Equation 3.1. S_1 and S_2 are top senders of these two ASes and the traffic demand can be represented by matrix T, specifying the traffic intensity from S_1 to AS_0 and AS_1, and S_2 to AS_0 and AS_1 respectively.
1) traffic matrix example 1: Consider this traffic matrix:

$$T_1 = \begin{pmatrix} 20 & 20 \\ 40 & 30 \end{pmatrix}.$$

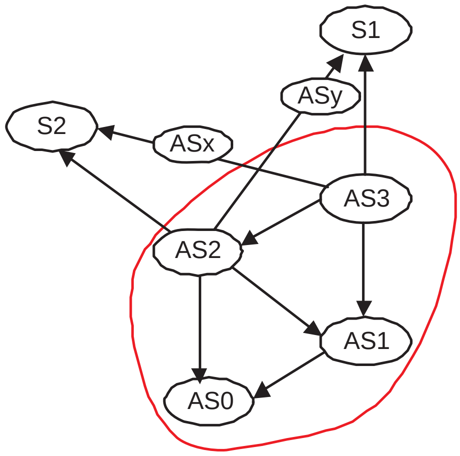

Figure 3.5: Example Network to Study Routing Convergence

The state transitions when multiple ASes are performing the Greedy ASPP Search Algorithm for their local load balance are summarized in Table 3.2. In the table, we also present our performance measures and the traffic intensity on each link in the network after each prepending action.

The initial load (*i.e.*, without prepending) on each link[2] is computed in the second row, based on applying Equation 3.2 to the shortest paths from sources to destinations before prepending. The ensuing *prediction* step shows the local evaluation by AS_0 and AS_1 of what will happen if a prepending action is taken. The subsequent *update* step shows what prepending actually takes place - only AS_0 decides to prepend the link $(2, 0)$. This action leads to new traffic load which may be different than the prediction if someone else also takes a prepending action (*e.g.* in AS_1's case). Then the two ASes perform prediction again, and this time neither finds any reason to prepend again. Both the global load balance index and scalability bottleneck improve slightly, at a moderate increase of total resource consumption A

[2]Note, (2,0) and (1,0) are provider links for AS_0; (2,1) and (3,1) are provider links for AS_1; (3,2) is a provider link for AS_2 but AS_2 is not doing load balancing since it has only one provider link.

(from 126 to 150). This is a desirable situation, as it shows that the decentralized load balancing actions converge and result in a reasonable global state.[3]

$\mathcal{P}$	AS_0			AS_1			AS_2	A	$\mathcal{G}_{lb}$	B
	$t_{(1,0)}$	$t_{(2,0)}$	$\mathcal{I}_{lb}(0)$	$t_{(2,1)}$	$t_{(3,1)}$	$\mathcal{I}_{lb}(1)$	$t_{(3,2)}$			
$\varnothing$	6	52	0.61	30	26	0.99	6	126	0.33	0.52
Prediction:										
(2,0)	40	20	0.90							
(2,1)				15	41	0.82				
Update:										
(2,0)	40	20	0.90	50	40	0.98	0	150	0.36	0.50
Prediction:										
(1,0)	6	52	0.61							
(2,1)				15	55	0.75				
no more prepending based on local predictions $\Rightarrow$ convergence.										

Table 3.2: The Network Routing Converges.

2) traffic matrix example 2: Consider a second example:

$$T_2 = \begin{pmatrix} 20 & 30 \\ 10 & 80 \end{pmatrix}.$$

When there is no prepending in the network, the traffic from S_1 to AS_0 goes through three paths (S_1, AS_y, AS_2, AS_0), (S_1, AS_3, AS_2, AS_0) and (S_1, AS_3, AS_1, AS_0). The traffic from S_2 to AS_0 goes through (S_2, AS_2, AS_0). Thus $t_{(2,0)} = 2/3 * 20 + 10 = 23.33$. Similarly, we can calculate the traffic on other links and predict what will happen in this network.

In Table 3.3, we do not show the prediction step in order to simplify the presentation. The first column shows the prepending *changes* (add or remove) because of the last execution of the Greedy ASPP Search Algorithm. One can find that the

[3] In this example, as well as the rest of examples based on the same network, we compute the global metrics based on only the four ASes and links between them. We ignore S_1, S_2 and other ASes introduced solely for forwarding traffic.

$\mathcal{P}$	AS_0			AS_1			AS_2
	$t_{(1,0)}$	$t_{(2,0)}$	$\mathcal{I}_{lb}(0)$	$t_{(2,1)}$	$t_{(3,1)}$	$\mathcal{I}_{lb}(1)$	$t_{(3,2)}$
$\varnothing$	6.67	23.33	0.76	80.00	36.67	0.88	6.67
AS_0 finds nothing to do. AS_1 finds that $(2,1)^1$ can improve its local metric.							
$(2,1)^1$	6.67	23.33	0.76	40.00	76.67	0.91	6.67
AS_0 finds that $(2,0)^1$ can improve its local metric. AS_1 finds nothing to do.							
$(2,0)^1$	20.00	10.00	0.90	40.00	90.00	0.87	0.00
AS_0 finds nothing to do. AS_1 finds that $(2,1)^{-1}$ can improve its local metric.							
$(2,1)^{-1}$	25.00	5.00	0.69	85.00	50.00	0.93	0.00
AS_0 finds that $(2,0)^{-1}$ can improve its local metric. AS_1 finds nothing to do.							
$(2,0)^{-1}$	6.67	23.33	0.76	80.00	36.67	0.88	6.67
The network goes back to the initial state.							

Table 3.3: Routing Instability Caused by Interference of ASPP.

prepending actions of AS_1 and AS_0 are interfering with each other. In this case, the prepending actions on links (AS_2, AS_1) and (AS_2, AS_0) are nullified which leads the network back to the original routing. The reason for this instability is that there is no solution for both ASes to balance their load at the same time, which means these ASes have conflicting prepending requirements. From the game theory point of view, we can say there is no *Nash Equilibrium* for this game played by AS_0 and AS_1. If there is no mechanism to stop it, neither of them would give up. Then the best routes for AS_0 and AS_1 involve an oscillation, and the network cannot reach a stable routing solution. In fact, in this example, the prepending policy is dependent on the traffic distribution and *vice versa*, thus the instability is similar with the instability caused by load-dependent routing.

We make the following observation from this example:

Observation 6 *If there are multiple ISPs in the network using ASPP approach for their traffic engineering goals, it may cause routing instability.*

3) traffic matrix example 3: Finally, we observe that the decentralized prepending actions may actually lead to a *"worse"* global state. Consider the following traffic matrix:

$$T_3 = \begin{pmatrix} 40 & 20 \\ 40 & 50 \end{pmatrix}.$$

In this case, AS_0 sees that one prepending on the link (AS_2, AS_0) can help balance its incoming traffic on links (AS_1, AS_0) and (AS_2, AS_0). After implementing this prepending, we observe all the global metrics are slight worse off, as shown in Table 3.4.

$\mathcal{P}$	AS_0			AS_1			AS_2	A	$\mathcal{G}_{lb}$	B
	$t_{(1,0)}$	$t_{(2,0)}$	$\mathcal{I}_{lb}(0)$	$t_{(2,1)}$	$t_{(3,1)}$	$\mathcal{I}_{lb}(1)$	$t_{(3,2)}$			
$\varnothing$	13	66	0.68	50	33	0.95	13	188	0.36	66
(2,0)	60	20	0.80	70	60	0.99	0	210	0.35	70
Neither AS_0 nor AS_1 finds useful prepending action.										

Table 3.4: Convergence to Worse Global State.

In this case, AS_0's prepending action leads to a better local load balance for both AS_0 and AS_1, yet globally, the metrics B and $\mathcal{G}_{lb}$ are slightly worse off. The reason is that the link (AS_3, AS_2) carries no traffic after the prepending. In this example, the global state is not much worse off than before, but the important lesson is that it can get worse after prepending. From this example, we make the following observation.

Observation 7 *The global state may be worse off when ASes are optimizing their individual local goals using ASPP approach.*

3.5 Finding ASPP Instability in the Internet

As we have stated, ASPP is one traffic engineering method achieved by explicitly announcing routes with inflated AS paths to influence other ISPs, thus it is possible to analyze its use based on publicly available routing tables. Route Views [39] operates a number of BGP data collection points that peer with

BGP routers at various ISPs. It captures snapshot every four hours from November 8th, 1997, containing more than 7,000,000 routes for more than 160,000 prefixes in each snapshot. This large database of routing tables presents a partial view of all the routes advertised by different ISPs over a period of ten years, which makes it possible to study the dynamic ASPP behavior of ISPs in the Internet.

One important observation in Section 3.4 is that ISPs may have conflicting requirements which result in *repetitive* adjustment of ASPP configurations. In this study, we analyze the routing information from the real-life Internet to see if the instability problem really occurs. From our analysis so far, indeed we find some interesting examples.

3.5.1 Finding Potential Conflicts

Let p denote a prefix, and R_p denote the set of all the routes to p stored in a routing table. For example, there is a prefix p, and

$$R_p = \{r_1,\ r_2\}$$

where

$$r_1 = (0, 1, 3, 3, 3, 4, 5),\ r_2 = (0, 2, 3, 5, 5, 5).$$

In this example, r_2 indicates that AS_5 prepends the route twice when it announces the route to AS_3, thus AS_5 appears three times in r_2. On the other hand, r_1 indicates AS_5 announces the route to AS_4 without prepending. One can infer that this implies AS_5 wants the traffic destined to p to go through AS_4 instead of AS_3. So we say AS_5 *prefers* r_1. Similarly, one can infer that AS_3 *prefers* r_2 because AS_3 wants the traffic destined to p to go through r_2. Obviously, this disagreement in preferences may cause the best route for p from some traffic sources to oscillate.

Formally, we define a function $l(i, r)$, to capture AS_i's prepending action on route r. Namely, $l(i, r)$ equals the number of times AS_i appears in route r.

Let A_p to be the set of all ASes that appear in at least one of the routes to p visible at Route Views, namely

$$A_p = \{i | \exists r \in R_p, l(i,r) > 0\}.$$

Based on $l(i,r)$ and A_p, we say AS_i *prefers* route r if any of following conditions is true:

1. $l(i,r) = 1$, *i.e.* AS_i is not doing any prepending to persuade others to not use r;
2. $l(i,r) = 0$, *i.e.* AS_i does not appear in r hence we assume it does not dislike it;
3. $l(i,r) > 1$ but it is the minimum number of appearance of i in any route $r \in R_p$.

Let $L(i,p)$ denote the set of routes for destination p that AS_i prefers. Then we say a prefix p may potentially have *conflicting routing requirements* (CRR) if:

$$\bigcap_i L(i,p) = \varnothing.$$

Let us go back to the earlier example. We know

$$A_p = \{0, 1, 2, 3, 4, 5\},$$

and

$$l(0,r_1) = 1, l(0,r_2) = 1, l(1,r_1) = 1, l(1,r_2) = 0,$$

$$l(2,r_1) = 0, l(2,r_2) = 1, l(3,r_1) = 3, l(3,r_2) = 1,$$

$$l(4,r_1) = 1, l(4,r_2) = 0, l(5,r_1) = 1, l(5,r_2) = 3.$$

Therefore,

$$L(0,p) = L(1,p) = L(2,p) = L(4,p) = \{r1, r2\},$$

$$L(3,p) = \{r2\}, \quad L(5,p) = \{r1\},$$

$$\bigcap_{i\in[0,5]} L(i,p) = \varnothing.$$

We see that the routes for this prefix satisfy the above CRR condition. There might be a disagreement between AS_3 and AS_5 about how to route packets to the prefix.

3.5.2 A Pathological Example

Based on Route Views database, we analyze how prepending configurations change over time using a total number of 388 snapshots, every four hours from 8pm February 24th, 2004 to 8pm April 30th, 2004.

Since there are too many prefixes in the database, the first thing we do is to find the prefixes that are highly likely to have conflicting prepending requirements. Based on a random picked snapshot, we extract 229 prefixes whose routes satisfy the CRR condition for further analysis.

For each extracted prefix, we analyze its routes over a period of several months to see how different ISPs change their ASPP actions. We also look into the corresponding routing tables to find the best route and see how the prepending actions affect the best route selection.

Our preliminary study reveals that some prependings appear only briefly. However, some other prependings *repeatedly change* during the whole period, which indicates likely interference between ASPP actions by different ISPs.

Table 3.5 shows detailed information for one pathological example, where the prefix $\mathcal{P}$ is 80.96.218.0/24, including the prepending actions and the best route. The first column shows the snapshot date (empty means the same date as the previous entry); the second column indicates the AS that performed the ASPP action; the second and the third column together give the link involved in the prepending; and the 4th and the 5th column show the change of the prepending length (from and to). The last column is the best route to prefix $\mathcal{P}$ at that time. For example, the 1st row means that AS_3 changed the prepending length from 3 to 0 when it announced routes to AS_5 during the period between 8am and 12am on March 15th 2004.[4]

date	AS_a	AS_b	l_{before}	l_{now}	*best route*
03-15-12	AS_3	AS_5	3	0	$S\ AS_5\ AS_3\ AS_1$
	AS_1	AS_2	2	3	
03-17-00	AS_3	AS_5	0	3	$S\ AS_4\ AS_2\ AS_1\ AS_1\ AS_1$
	AS_1	AS_2	3	2	
03-26-08	AS_3	AS_5	3	0	$S\ AS_5\ AS_3\ AS_1$
	AS_1	AS_2	2	3	
03-26-12	AS_3	AS_5	0	3	$S\ AS_4\ AS_2\ AS_1\ AS_1\ AS_1$
	AS_1	AS_2	3	2	
03-29-12	AS_3	AS_5	3	0	$S\ AS_5\ AS_3\ AS_1$
	AS_1	AS_2	2	3	

Table 3.5: Prepending Changes in Routes to Prefix $\mathcal{P}$

From Table 3.5, one can find the prepending lengths on the links (AS_5, AS_3) and (AS_2, AS_1) alternatingly change during the period between 12pm March 15th, 2004 and 4am March 28th, 2004. We observe AS_3's prepending length changes from 0 to 3 and back many times, and AS_1's prepending length changes from 2 to 3 back and forth in a similar way as the example shown in Table 3.3. As a result of these repetitive prepending actions, the best route for this prefix involves an oscillation.

In order to clarify the situation, we infer a simplified topology graph, showing all routes (over forty) from Route Views to the prefix $\mathcal{P}$ for each snapshot. The first two snapshots are shown in Figure 3.6. The 3rd and 5th snapshots are the same as the first one, and the 4th and 6th snapshots are the same as the second one. We see that there are only two groups of routes from Route Views to this prefix: one group of routes go

[4]Note, we have replaced the real AS number with shorthand here.

through the link (AS_2, AS_1), and the other group of routes go through the link (AS_5, AS_3). The best route, from the Route Views vantage point, actually *oscillates* between the two subpaths $(\ldots AS_4, AS_2, AS_1)$ and $(\ldots AS_5, AS_3, AS_1)$, depending on the relative amount of prepending applied by AS_1 on (AS_2, AS_1) and by AS_3 on (AS_5, AS_3).

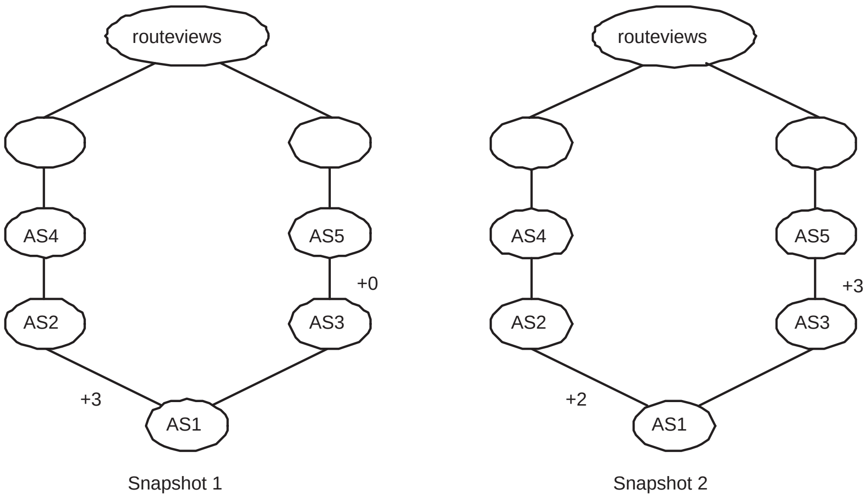

Figure 3.6: Routes to Prefix $\mathcal{P}$ from Route Views

A possible explanation of what is going on is as follows. At the left snapshot, the best route goes through AS_5, thus AS_3 wants to increase the prepending length (from 0 to 3) to reduce the traffic it transits, while AS_1 wants to decrease the prepending length to induce more traffic on the link from AS_2. Because of these two prepending changes, the best route becomes the path through AS_2 as shown in the second snapshot. Since there is too much traffic shifted from the right path to the left path, now AS_3 wants to increase the traffic on $(AS_5, AS3)$, while AS_1 wants to decrease the traffic on (AS_2, AS_1). Then both of them decide to revert their earlier changes, and the oscillation occurs.

Note that there are some limitations with using Route Views for our analysis. Route Views receives only the best route for each prefix from all its neighbors. Therefore, we can only catch the situation when the route with changing prepending is also the best route for neighbors of Route Views. Clearly this only gives us a small subset of all the potential cases.

Changes in prependings may also be caused by other reasons, *e.g.*, simply due to the non-stationarity of traffic. We have already carried out some analysis of prepending changes for random prefixes found in the Route Views database. Indeed they change at a much lower rate.

3.6 Guidelines to Avoid ASPP Instability

Routing instability can result in poor end-to-end network performance and degrade the overall efficiency of the Internet infrastructure. In this section, we present some guidelines for network operators to avoid the routing instability caused by distributed prepending actions. In these guideline, we assume that all ASes are deploying the popular routing policy, *i.e.*, selective announcement and typical local preference.

Guideline 1 *If only stub ASes are performing prepending on their local provider links, these prepending actions will not result in routing instability.*

This has been stated as Observation 5. Note that this guideline does not allow transit ASes to perform any prepending, hence it is quite restrictive. A less restrictive guideline is as follows:

Guideline 2 *If no AS performs prepending except on the routes originated by itself, these prepending actions will not result in routing instability.*

Proof: In order to prove this guideline, we want to put all ASes in different "levels". Let us classify all ASes into different "*levels*". Let V_0 be the lowest level such that it contains all stub ASes only. Let V_1, $V_2 \ldots$ be the successively higher levels. An AS v belongs to V_i if all its customers come from levels V_j, where $j < i$.

Under Guideline 2, prepending actions of an AS can only affect traffic destined to itself. This traffic will not traverse any inbound provider links of a lower or same level AS because all AS paths should be *"valley free"* based on the *selective announcement* export policy. Therefore, any prepending by an AS will not affect the prepending decision of another AS at the same level or a lower level, even under load-dependent AS prepending scenario. Let us call this property 1.

Now we can prove the convergence by induction. Suppose at some time t_i, all ASes in levels from 0 to i have completed their prependings and will not do any more prepending. Therefore, ASes in level $(i + 1)$ will do their prependings without being affected by any prepending actions from these lower level ASes. Based on property 1, we know prepending actions from ASes at the same or higher level cannot affect the prepending decisions of ASes in this level $i + 1$ either. So there must exist some time t_{i+1} so that all ASes in levels from 0 to $i+1$ will have completed their prependings and will not change any more.

Again from property 1, we know ASes in level 0 cannot be affected by any prepending actions of ASes from the same or higher level. Therefore, there must exist a time t_0 that they complete their prepending without further change. Since there are finite number of levels in the Internet, the whole process must converge. This completes our proof by induction. ∎

Basically, we propose that ASes should not do prepending on the transit traffic. In practice, transit ASes may lose business (*i.e.* transit traffic) due to their prepending actions. Since they would like to induce more transit traffic to make more money, ASPP is not a suitable approach for them to do traffic engineering to some extent. However, the measurement study in [18] shows many transit ASes are performing ASPP, hence we present the following relaxed guideline:

Guideline 3 *If every prefix in the network has only one owner, and only the owner can do prepending on routes to the prefix, the distributed prepending actions will not result in routing instability.*

We suggest that one AS must announce its ownership of the prefix before it uses the traffic to this prefix for traffic engineering, and the AS can announce its ownership only when the prefix does not have an owner. In this way, only one AS may prepend this route. Therefore, we can follow the proof of guideline 2 and prove this new relaxed guideline. Under this guideline, ASes can do prepending on their transit traffic if downstream ASes do not announce ownership for the prefixes. Clearly, ASes near the origin AS should have a higher priority than ASes far away from the origin AS to be the owner of the prefix, since provider ISPs should respect their customer ISPs.

Here, we assume transit ASes are able to do prefix-based prepending, while the first two guidelines focus on AS-based prepending and destination-based prepending. Our measurement study in [18] shows more than 65% multihomed transit ASes are deploying prefix-based prepending currently. This makes the third guideline feasible in the Internet.

In order to implement this guideline, we can define some special "community" attribute values for global coordination in ASPP practice. An AS needs to signal its use of certain prefixes for traffic engineering by associating them with the designated community attribute value. This then prevents other ASes from using the same prefix for traffic engineering. However, the global distribution and synchronization of those community attribute are still problems.

3.7 Related Work

Traffic engineering refers to the process of tuning routing policies for a target objective. This process can be summarized into the following steps. Suppose an operator realizes that a particular outgoing edge link is congested. The first step is *traffic analysis* - fine-grain measurement data (such as Netflow) can be used to identify the destination prefixes responsible for the bulk of the traffic traversing this link; historical measurement data could be used to determine which of these prefixes have stable traffic volumes. The next step is *routing analysis*. For example, the operator could analyze the routing data to focus on the popular, stable prefixes that have a single "best" AS path across all of the egress points, in order to reduce the impact on neighboring ASes. Then, the operator needs to propose a suitable *routing policy configuration*. For example, the operator could consider modifying the import policy at the congested router to assign a lower local preference to some of these destination prefixes to divert this traffic to other egress links. The final step is *evaluation or test*, where the operator could use the prediction tool to check whether the proposed policy change can help the AS achieve the target objective.

In [32], the authors state that traffic engineering tools could ultimately evolve to automate these steps. Recently, a few companies have implemented solutions that can attach appropriate values of the local preference attribute for multihomed stub ASes to engineer their interdomain traffic based on traffic measurement and active performance measurement. However, a widespread deployment of such tools would heavily increase the possibility of BGP routing instability. Any dynamic interdomain traffic engineering technique, especially automated process, should be studied carefully before being deployed.

Based on BGP routing tables from routers connected to the AT&T backbone, Feamster *et al.* report that over 30% of the routes have some amount of ASPP, and most of these paths are prepended with one or two ASes [31]. However, based on routing tables from Route Views, the measurement in [18] shows that the share of prepended routes was only 7% in 1997 and it has been increased to more than 12% in 2004. This significant difference shows that it is important to observe ASPP on different levels of the Internet routing hierarchy.

In [25], Swinnen *et al.* conduct simulations to evaluate the ASPP method. In the simulation model, each stub AS is connected to two different transit ASes. When each stub AS conducts prepending on one of its provider link, the simulation results show that the distribution of the interdomain paths changes for almost all stub ASes. With a prepending length of 2, almost all the interdomain paths are shifted to the nonprepended links. Beijnum also studies the impact of ASPP on a doubly homed stub AS under two different scenarios in [15]. The first one is when the stub AS is doubly homed to similar ISPs in the sense that the ISPs directly peer with each other via the network access point. The second case is when the stub AS is doubly homed to dissimilar ISPs that do not directly peer with each other. He presents a simple example to show that applying ASPP to the second case has a more gradual effect on the change of the incoming traffic distribution.

Lo *et al.* conduct an active measurement in RIPE NCC network to study the route-level effects of prependings [43]. They inject beacon prefixes and change the AS path prepending length of those beacon prefixes every 2 hours for 26 hours. The results reveal a number of hidden processes in the course of propagating prepended routes, which is useful for explaining the method's efficacy and for systemizing the often *ad hoc* prepending procedure.

Motivated by a lack of systematic procedure to tune the ASPP, Chang and Lo propose a procedure to predict the traffic change before effecting the tuning. They implement and test the procedure in an operational, doubly homed AS which is connected to two regional ISPs [38]. The measurement results show that the prediction algorithm is fairly accurate. Moreover, the traffic shift peaks when the prepending length is changed from 2 to 3, and almost 60% of the routes are affected.

In [41], the authors propose a polynomial-time algorithm that determines the optimal prepending length vector for an advertised route at each ingress link of the target network. Specifically, given a set of elephant source networks and some maximum load constraints on the ingress links of the target AS, their algorithm determines the minimum prepending length at each ingress link so that the load constraints are met, when it is feasible to do so. However, their algorithm requires as input an AS Path length estimation from each source network to each ingress link. The whole algorithm relies on accurate parameters that can be only roughly estimated in practice. As a result, it only provides the best-case scenario for the effectiveness of prepending if all the required information is available. Although they also develop a robust variation (RPV) of that algorithm to deal with unavoidable inaccuracies in the AS Path length estimates, and to compensate for the generally unknown BGP tie-breaking process in upstream networks, the algorithm cannot work well when there are tight maximum load constraints.

In [13], the authors show how to compute optimal prepending both by using an Integer Linear Programming formulation and by exploiting Computational Geometry techniques. However, their model also requires that the lengths of the shortest AS-paths from each AS to each upstream are known. And they assume symmetric routing for those critical ASes not covered by their data sources. These two assumptions may bring important problems to the practical impacts of their approach. The algorithm we proposed in this chapter does not require the path length information as input, which makes it more attractive for ISPs.

The routing instability induced by policy disputes is first presented by Kannan Varadhan *et al.* in [21]. The route instability problem with using MED attribute has been studied in [52]. However, all of these theoretic analysis are based on abstract routing formalisms, such as Simple Path Vector Protocol (SPVP) and Stable Path Problem. Moreover, some analysis is only for a very restricted class of interdomain topologies. As far as we know, we are the first to pinpoint the essential feature of policy configuration for traffic engineering and present the routing instability problem caused by distributed prepending actions.

3.8 Summary

In this chapter, we focus on the ASPP approach to study various issues in interdomain traffic engineering. Although this approach has been practiced by many AS operators for a long time, there still lacks a systematic study of the approach and understanding of its effects and performance implications. As the ASPP approach continues to be applied by more and more ASes in the Internet, its effectiveness and potential problems should be studied in detail.

We first introduce the basic concept and applications of the ASPP approach in today's Internet. In particular, we summarize some justifications of performing ASPP in real world based on discussions with NANOG subscribers. Then we present a greedy algorithm to help ISPs perform ASPP systematically and efficiently. Our simulations demonstrate the performance of this algorithm.

As we know, selfish behavior of nodes in a distributed system may cause the routing instability problem and performance degradation. The decentralized interdomain traffic engineering practice by ISPs may bring the same issues to the Internet. Based on our network model and performance measures, we present various interactions of ASes when they use the prepending approach to perform their local optimization. We demonstrate that these local optimization behavior can result in routing instability and global performance degradation due to the lack of global coordination.

From our measurement study of ISPs' dynamic prepending behavior based on routing tables from Route Views, we observe that pathologic instability cases really happen in the Internet although the reason is not clear. We also present some simple guidelines for ISPs to perform ASPP properly to avoid the instability problem.

We have made several contributions towards using the ASPP approach to do inbound traffic engineering. Our modelling and analysis take a systematic look at the fundamental issues: (a)how is AS path prepending done locally? (b) what are important global metrics? (c) does the decentralized ASPP process always converge? (d) if it does, does it improve global metrics? (e) if it does not ensure convergence, how to avoid routing instability? We show that there is no interference between local prepending actions when only stub ASes do load balancing, and expose some problems (convergence, optimality) when providers are also doing load balancing.

The routing instability caused by policy disagreements has been studied by some researchers before, but the essential features of policy configurations for interdomain traffic engineering have not been pinpointed, so that they do not analyze the deep reasons for which the policy configurations in their examples appear, which is very important to connect practical problems with the theoretical analysis. We believe our analysis will help network operators to perform traffic engineering properly and avoid possible pitfalls.

☐ **End of chapter.**

Chapter 4

Economic Analysis of Overlay Streaming

Summary

Traditionally, how traffic flows through the Internet is determined by ISPs through the policy-based routing, where transit relationship is the primary factor. The advent of overlay networks, however, changes the rules by providing traffic favoring applications' needs. In this chapter, we capture the important difference between these two routing mechanisms by modeling the routing in the overlay streaming network as to optimize the performance of transferring each byte of its subscribers' demand. Then we conduct an economic analysis for a marketplace with two competing ISPs to illustrate the economic implications of the routing tussle. Our analysis reveals that overlay networks can result in the "fairness" problem in distributing the free peering benefits. ISPs must control their provider link capacity as well as free-peering link capacity they provide to their peering ISPs to sustain their relative economic positions despite of application layer routing.

ISPs rely on BGP, a policy-based routing protocol to control the routing of interdomain traffic according to their individual business concerns. For example, in Figure 4.1, ISP_C is a transit provider, and ISP_A and ISP_B are customers of ISP_C. At the same time, ISP_A and ISP_B set up a free peering link to reduce the transmission costs on their links to ISP_C. In this network, if a subscriber i of ISP_A needs an object r from the Internet, the information flow can transit through ISP_C and arrive at ISP_A for subscriber i. Due to the control of policy-based routing by ISPs, the object cannot traverse the path (Internet $\ldots ISP_C, ISP_B, ISP_A$) since ISP_A cannot receive the route to r from ISP_B under the free peering agreement.

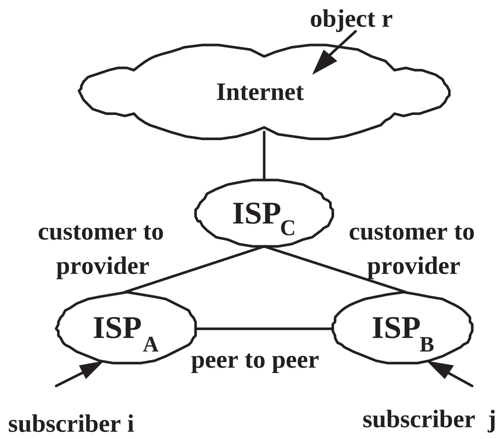

Figure 4.1: Example to Illustrate Routing Control.

It is reasonable to assume that the routing culminated from such decentralized peering agreements is *economically efficient*. If ISP_A drastically over-provisions its provider link, another competing ISP (*i.e.* ISP_B) with a lower operating cost would be in the position to undercut the price ISP_A charges its subscribers thus grab customers away from ISP_A. Similarly, if ISP_A and ISP_B do not freely exchange their local traffic, other competing ISPs may do so to undercut their business.

Such economic efficiency, however, does not imply best possible service for individual subscribers and the applications they are running. This conflict may due to a variety of reasons. For example, one reason is that it would be too complicated for the network layer to learn about all the application requirements and tailor its routing accordingly; another reason is that applications' routing preference might be in conflict with ISPs' business consideration.

An important class of applications that come into conflict with ISP controlled network routing is the peer-to-peer (P2P) applications. To provide efficient and speedy distribution of content to many receivers, peers play the role of information receiver as well as server at the same time. It causes information to be routed according to the distribution of the overlay nodes (peers), and to some extent the performance of paths between the peers. The resultant patterns of information flow may be quite different from that intended by ISPs' transit agreements. Imagine the case that the object r is a P2P object in the Internet and it is needed by both subscriber i and subscriber j. To improve performance and scalability, a P2P application makes both subscriber i and j to provide service to each other. So subscriber i may receive some pieces of r from subscriber j and vice versa. If we look at the routing of the pieces from subscriber j to subscriber i at the application layer, they traverse along the path (Internet, ISP_C, ISP_B, ISP_A). This part of traffic on the peering link is beneficial only for ISP_A. In this sense, ISP_B is providing transit service for ISP_A without being paid by ISP_A, which is not the intention of the peering agreement between ISP_A and ISP_B. In this situation, policy-based routing (BGP) fails to implement the intended selective transit service because the path is implemented through two network layer connections and both connections are legitimate under the traditional free peering agreement. This example illustrates the routing tussle between ISPs and P2P applications and it shows that policy-based routing cannot always control the routing of P2P traffic effectively. Moreover, ISP_A may also provide transit service for ISP_B without being paid since subscriber j may also receive some pieces of r from subscriber i. So the question becomes which ISP receives more benefit from this free peering, and whether it is fair.

In many networks, the overlay traffic, especially P2P traffic, has already overtaken the traffic volume generated by traditional high-volume applications such as web and email. For this reason, the routing tussle between ISPs and overlay networks has become a significant problem for ISPs' business model [1]. This problem also bears significance on Internet's service model, and possible future requirements for policy-based interdomain routing protocols.

In this chapter, we will study this problem based on a simple optimization routing model and provide some insights on the economic implications of overlay networks. We conclude that ISPs must control their provider link capacities as well as free-peering capacities they provide to other peers to maintain certain parity between different incoming link capacities among all ISPs. Only when such parity is maintained, ISPs can sustain their relative economic positions despite of application layer routing.

4.1 The Market under Study

We focus on the local market shown in Figure 4.2 to study the economic implications of various routing mechanisms on ISPs' peering strategy. The market is with n homogeneous subscribers, and they can access the Internet via two ISPs, say ISP_1 and ISP_2. We view each local ISP to be simply in the

[1] We have learned about the problem from talking to ISPs in our region.

business of providing transit service, and compete with other ISPs in similar positions for subscriber share. We denote ISP_i's market share as α_i $(i = 1, 2)$, and $\alpha_1 + \alpha_2 = 1$. Without loss of generality, we assume $\alpha_1 \geq \frac{1}{2}$ in the rest of this chapter.

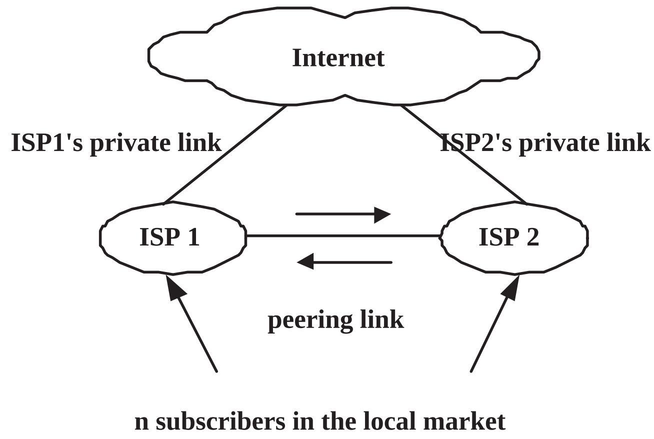

Figure 4.2: The Local Market under Study.

Both ISPs need to set up connections with their respective providers (which could be the same) for access to the rest of the Internet. In practice, many ISPs are multihomed, which means they have multiple provider links to reach the Internet. In our analysis, we use one virtual link between ISP_i and the Internet to denote these multiple physical links and this virtual link is referred to as ISP_i's *private link*. These two local ISPs may also set up a free *peering link* between them to reduce traffic on their respective private links. The traffic exchange on the peering link is free of charge. In addition, we ignore the peering cost such as maintenance cost in this analysis.

In the following discussion, the incoming capacity of ISP_i's private link (from the Internet to ISP_i) is denoted by c_i^o, and the traffic intensity on ISP_i's private link is denoted by t_i^o. Similarly, the incoming capacity of ISP_i's peering link (from ISP_j to ISP_i) is denoted by c_i^p, and the traffic intensity on that link is denoted by t_i^p. When $c_1^p = c_2^p = 0$, it means that there is no peering link between two ISPs.

4.2 Methodology for Economic Analysis

Obviously, each subscriber needs to pay its provider for the access service. As it is quite common with current ISP pricing, we assume that each ISP uses flat rate charging and charges a fixed price p_i, and it is possible that $p_1 \neq p_2$.

ISPs need to pay their providers for the traffic exchange on their private links. In real life, the cost structure for an ISP can be quite complicated. The dependency on traffic is often not exact, but based on some measured percentile of traffic intensity over a period of time. There is usually also a cost associated with the capacity and guaranteed traffic minimum. In order to simplify the analysis, we assume a constant cost q per unit intensity of traffic and this cost is the same for all ISPs in the same local market.

The net income of ISP_i is therefore

$$\mathcal{R}_i = p_i n \alpha_i - q t_i^o,$$

wherein $p_i n \alpha_i$ is the revenue from its subscribers and $q t_i^o$ is the transmission cost. If $\mathcal{R}_i < 0$, ISP_i cannot survive economically. We define the minimum price that ISP_i needs to charge its subscribers in order to survive as ISP_i's *break-even price*, which is denoted as p_i^*. From the last equation, ISP_i's break-even price is given by

$$p_i^* = \frac{q}{n \alpha_i} t_i^o. \tag{4.1}$$

If one ISP's break-even price becomes lower after peering, we say it benefits from peering. We use the decrease in the break-even price to measure the benefit an ISP achieves from the peering. It is also an ISP's concern to solidify its market position by driving towards a lower break-even price than its competitors. In the two-ISP model, ISP_1 would like to minimize $p_1^* - p_2^*$, while ISP_2 would like to minimize $p_2^* - p_1^*$ in order to be more competitive in the market.

In the rest of this chapter, we use p_i^* to denote the break-even price of ISP_i after peering with limited capacity. In particular, $p_i^*(0)$ is the break-even price of ISP_i under no peering scenario; and $p_i^*(\infty)$ is the break-even price under unlimited capacity peering scenario.

4.3 Traffic and Routing

The traffic demand can be defined in terms of end-to-end flows. Specifically, *local traffic* is defined as the traffic between two subscribers in one ISP network, *peering traffic* is defined as the traffic between subscribers in different ISP networks in the local market; whereas *remote traffic* is defined as the traffic between local subscribers and non-local subscribers in the Internet.

Conventionally, only peering traffic is allowed to be exchanged through the peering link and this traffic exchange is beneficial to both ISPs. A previous paper [28] studied ISP peering and provided a good understanding of peering strategies based on a network model with only local traffic and peering traffic. In this analysis, we focus on the effects of remote traffic on local peering decisions, which is the new problem coming with the appearance of overlay networks. In particular, we focus on *incoming* remote traffic generated by local subscribers who want to download objects from the Internet. In local markets, incoming remote traffic tends to dominate outgoing remote traffic and the normal settlement model for provider links is based on the maximum of the two directional traffic. So this assumption is reasonable to simplify the analysis.

In order to study the economic implications of the traffic brought by new application layer routing, we assume there is an object steaming with intensity λ in the Internet, and this steaming is needed by subscribers in both ISPs. The distribution of this resource to each ISP is done by an overlay, then the ISP can distribute the resource to its subscribers using its own network. This can be viewed as a simplified model of a CDN overlay network. Under this situation, each ISP has a interdomain demand of λ, whose ISP level routes is determined by application layer routing of the overlay network.

4.3.1 Routing Model

In order to study the economic implications of overlay traffic on ISPs' peering strategies, we must model the routing behavior of the overlay network so that we can predict the traffic intensity on various links and analyze the economic consequences of the traffic.

First, let us discuss all possible ways for two ISPs to download a content object in the streaming.

1. Both ISPs download the content object via their own private links.

2. One ISP downloads the content object via its own private link, and then the other ISP downloads the object via the peering link.

Different applications/nodes may use different metrics and algorithms to optimize their routing decisions. The essence of the routing behavior, we believe, can be captured by an optimization model, as if the nodes in the overlay network try to optimize the performance of transferring each byte of their traffic.

We define the following partitions of the demand: x_i $(i = 1, 2)$ is the fraction that goes to ISP_i only; and y_i $(i = 1, 2)$ is the fraction that goes to ISP_i first and is subsequently downloaded

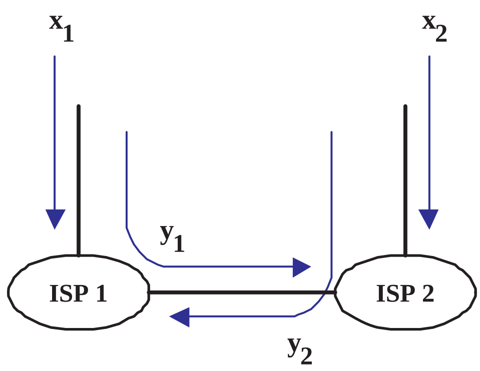

Figure 4.3: Routing Model.

by ISP_j ($j = 1, 2$ and $j \neq i$) through the peering link. Obviously, x_i and y_i are parameters that reflect the routing in this network. Figure 4.3 illustrates these parameters.

Therefore, the traffic intensity on ISP_i's private link is $t_i^o = \lambda(x_i + y_i)$ ($i = 1, 2$). The traffic on the peering link from ISP_j to ISP_i is $t_i^p = \lambda y_j$.

In our model, we define the goal of application layer routing as to determine x_i and y_i to optimize the performance of transferring each byte of its demand. Let $g(t, c)$ be some performance metric that measures the transferring performance on a link with capacity c when the traffic intensity on this link is t. Since the flows on different links affect different number of subscribers and different amount of traffic, we need to give the transferring performance of each link a weighting factor $w(\alpha, t)$, where α represents the number of subscribers affected and t denotes the traffic intensity over this link. Thus, we add up the weighted performance of all links in the local network and get the following expression. It is used to measure the performance of transferring all traffic in the whole network.

$$\begin{aligned} & w(\alpha_1, t_1^o) g(t_1^o, c_1^o) + w(\alpha_2, t_2^o) g(t_2^o, c_2^o) \\ + \; & w(\alpha_1, t_1^p) g(t_1^p, c_1^p) + w(\alpha_2, t_2^p) g(t_2^p, c_2^p). \end{aligned} \tag{4.2}$$

Overlay routing optimizes Equation (4.2) subject to the following demand constraint and capacity constraints:

$$\begin{array}{rl} \text{demand constraint} & x_i + y_1 + y_2 = 1, \quad (i = 1, 2) \\ \text{capacity constraints} & t_i^o = \lambda(x_i + y_i) \leq c_i^o, \\ & t_i^p = \lambda y_j \leq c_i^p. \end{array} \tag{4.3}$$

For this analysis, we assume

$$g(t, c) = c - t, \tag{4.4}$$

which is the available bandwidth of the link, to measure the performance of transferring each byte of the traffic on this link. The weighting factor is defined as

$$w(\alpha, t) = \alpha t, \tag{4.5}$$

where t is the traffic volume on the link and α is the percentage of subscribers involved. Roughly speaking, $\alpha t(c - t)$ can be interpreted as the sum of the performance of transferring each byte for all traffic and all related subscribers on the link. The overall routing objective of the overlay network is to optimize the transferring performance for the traffic on all different links in the network.

Furthermore, from Equation (4.3), we know that $x_i = 1 - y_1 - y_2$, which implies $x_1 = x_2$. It is obviously true according to our definition. However, it does not mean that two ISPs download the same amount of the demand since the portion of the demand ISP_i needs to download is $x_i + y_i$ and generally $y_1 \neq y_2$.

Let us use λ as the unit to measure the capacity or traffic intensity of different links, thus the intensity of the demand can be considered as 1. The last optimization problem can be refined to the following form:

$$\begin{array}{rl} \mathbf{min}_{y_1, y_2} & \alpha_2(y_1 - y_1^{'})^2 + \alpha_1(y_2 - y_2^{'})^2, \\ \textbf{subject to} & l_i \leq y_i \leq h_i, \\ & y_1 + y_2 \leq 1, \end{array} \tag{4.6}$$

where

$$
\begin{aligned}
y_1^{'} &= \frac{c_2^p - c_2^o + 2}{4}, \\
y_2^{'} &= \frac{c_1^p - c_1^o + 2}{4}, \\
h_i &= \min\ (c_j^p,\ 1), \\
l_i &= \max\ (1 - c_j^o,\ 0).
\end{aligned}
$$

4.3.2 Analysis of the Routing Solution

From Equation 4.6, we can see that the objective function is a convex function in a three-dimensional space. Let us define the rectangle $y_i \in [l_i, h_i]$ as the *feasible rectangle.* As shown in Figure 4.4, the feasible region is confined by the feasible rectangle and the line $y_1 + y_2 = 1$, indicated by the shaded area. Although it is difficult to express the solution by a clean closed form equation, where the optimal solution lands in the feasible region has clear physical meanings. We discuss different possible scenarios as follows [2] :

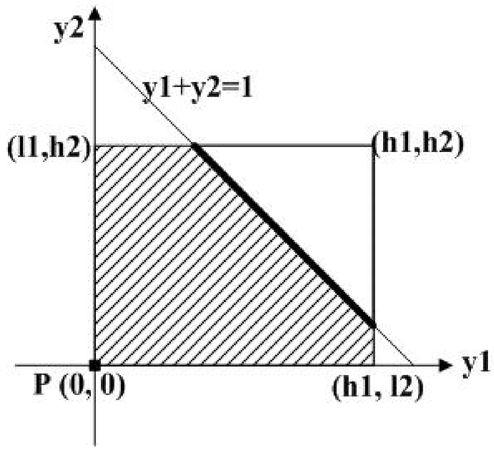

Figure 4.4: Feasible Region.

[2]Note the location of the rectangle and the line depends on the parameters. This figure is not exact and it is only used to illustrate different scenarios.

1. If $(y_1^{'}, y_2^{'})$ satisfies all constraints, we will have $(y_1^*, y_2^*) = (y_1^{'}, y_2^{'})$. In this case, the optimal solution is also the global optimal solution of the unconstrained problem. If the ISPs are cooperative, they can collectively adjust their link capacities to achieve this global optimal result.

2. If $y_1 + y_2 = 1$ is an active constraint, the solution lies on the bold line in the figure and it means all demand would traverse only one of the private links. Whatever is downloaded by ISP_i would then be downloaded by ISP_j $(j \neq i)$ over the peering link. This is achieved only when the performance of the peering link is good enough for the two ISPs to exchange demand, so we refer to this scenario as "*Unlimited Capacity Peering*".

3. If $y_1^* = y_2^* = 0$, which is indicated by the point P in the figure, it means there is no traffic exchange on the peering link. We refer to this scenario as "*No Peering*".

4. If $y_1 + y_2 = 1$ is not an active constraint and $y_1^* = y_2^* = 0$ is not true, the solution should be the point within the shaded rectangle and *closest* to $(y_1^{'}, y_2^{'})$. We also have $y_1^* + y_2^* \in (0, 1)$, which means the capacity of the peering link is not large enough so that two ISPs cannot share the demand efficiently. We refer to this scenario as "*Limited Capacity Peering*".

The routing model predicts the traffic intensity on various links under different ISP peering (*i.e.* c_1^p and c_2^p) and provisioning (*i.e.* c_1^o and c_2^o) agreements. In the following section, we will analyze the economic consequences of this traffic.

4.4 Economic Analysis of Different Scenarios

The methodology described in Section 4.2 can be readily applied to analyze the implications of the overlay streaming traffic on ISPs' peering. We first simplify the presentation of the optimization problem in Equation 4.6 as follows:

$$\begin{aligned}
\mathbf{min}_{y_1,y_2} \quad & \alpha_2(y_1 - \frac{c_2^p - c_2^o + 2}{4})^2 \\
& + \alpha_1(y_2 - \frac{c_1^p - c_1^o + 2}{4})^2, \qquad (4.7)\\
\textbf{subject to} \quad & max(0, 1 - c_j^o) \le y_i \le min(1, c_j^p), \\
& y_1 + y_2 \le 1.
\end{aligned}$$

Next, we will analyze these scenarios one by one: (a) *no peering*, (b) *unlimited capacity peering*, and (c) *limited capacity peering*.

4.4.1 No Peering Scenario

Recall that the routing solution $y_1^* = y_2^* = 0$ means that there is no exchange of traffic over the peering link. For the no peering scenario to occur, the network must satisfy the following two conditions[3]:

$$c_i^o \ge 1 \qquad i = 1, 2 \qquad (4.8)$$

$$c_j^p = 0 \quad \text{or} \quad c_i^o - c_i^p \ge 2 \qquad i, j = 1, 2 \quad i \ne j \qquad (4.9)$$

Equation 4.8 guarantees feasibility, namely each ISP can independently satisfy its interdomain demand respectively. Equation 4.9 means either there is no peering link, or the peering link has significantly lower capacity than the private links such that there is no local traffic exchange under optimal routing. In the no peering scenario, we can make the following observation about the economic positions of the two ISPs:

[3]The derivation of the conditions and the solutions for different scenarios can be found in the appendix.

Proposition 1 *In a network with dominant overlay streaming traffic, an ISP's break-even price is inversely correlated to its market share when there is no local peering. In other words, the ISP with the highest market share has the lowest break-even price.*

In economics jargon, we have the *economy of scale*, which means the ISP with more market share is more efficient in making use of its private link to satisfy its subscribers, and as a result has a more competitive market position.

Proof: From Equation 4.1, we have

$$p_i^*(0) = \frac{q}{n\alpha_i} t_i^o = \frac{q}{n\alpha_i}\lambda = \frac{q}{n\alpha_i}, \tag{4.10}$$

and observe that $p_i^*(0)$ is a decreasing function of α_i. ∎

Note that the statement above does not necessarily mean ISPs with higher market share prefer no peering. As analyzed in [28], as a way to provide short circuits for local traffic, local ISPs may choose to set up free peering to improve performance and reduce the cost of transiting local traffic over the private links, resulting in benefits for both ISPs. Now let us consider the possible effect of the local peering in a network with dominant overlay streaming traffic demand.

Proposition 2 *In a network with dominant overlay streaming traffic, local peering improves the overall efficiency of the peering ISPs, and each ISP's break-even price is always "better off" or "equal to" before.*

Proof: Let us consider two ISPs as one network. Before peering, the cost of this network is $q(\lambda + \lambda) = 2q\lambda$. After peering, the cost of this network is $q\lambda(x_1 + y_1 + x_2 + y_2) = q(1 + x_1)\lambda = q(1 + x_2)\lambda \leq 2q\lambda$. Although the local peering incurs some operating cost to the peering ISPs, it is usually not significant and can be ignored here. We see that the total cost of this network is reduced, so the overall efficiency of this network is better than before. Furthermore, since

$$p_i^* - p_i^*(0) = \frac{q}{n\alpha_i}(x_i + y_i - 1)\lambda = \frac{q}{n\alpha_i}(-y_j)\lambda \leq 0,$$

So the break-even price would be lowered (or at least equal to before) for both ISPs after peering. ∎

This proposition shows the free peering is still beneficial for the whole network with overlay streaming traffic. However, the proof also shows the break-even price reduction $(\frac{q}{n\alpha_i}y_j)$ for two ISPs might be different. Under certain conditions, we may have the routing solution $(y_1^*, y_2^*) = (1, 0)$. It means all the public demand would become traffic for ISP_1's private link, and ISP_2 always downloads the remote objects from ISP_1. Clearly, peering has not helped ISP_1 at all while ISP_2 takes full advantages of ISP_1's committed bandwidth without having to pay for it. In this extreme case, clearly there is no incentive at all for ISP_1 to peer with ISP_2.

This extreme case highlights some of the implications of application layer routing on ISPs' peering agreements. It gives an intuitive explanation of why the recent trend of increasing P2P traffic in ISP network is causing some ISPs to revise their peering relationships. In the following subsections, we will discuss the distribution of peering benefit in different scenarios, which can shed light on how an ISP should make peering decision.

4.4.2 Unlimited Capacity Peering Scenario

In this scenario, $y_1^* + y_2^* = 1$. In other words, each remote object downloaded by ISP_1 is shared with ISP_2 and vice versa. The routing problem in Equation 4.7 can be transformed to the following equivalent problem:

$$\begin{aligned} \mathbf{min}_{y_1} \quad & (y_1 - \frac{\alpha_2(c_2^p - c_2^o) - \alpha_1(c_1^p - c_1^o) + 2}{4})^2, \\ \textbf{subject to} \quad & y_{1l} \le y_1 \le y_{1h}, \end{aligned} \tag{4.11}$$

where y_{1l} and y_{1h} are x-coordinates of two intersection points of the line $y_1 + y_2 = 1$ and the rectangle in Figure 4.4. In this simplified form, the optimal solution can be easily derived:

$$y_1^* = \begin{cases} \frac{1}{2} + t & : \text{ if } t \in [y_{1l}, y_{1h}] \\ y_{1l} & : \text{ if } t < y_{1l} \\ y_{1h} & : \text{ if } t > y_{1h} \end{cases} \tag{4.12}$$

where $t = \dfrac{\alpha_2(c_2^p - c_2^o) - \alpha_1(c_1^p - c_1^o)}{4}$. The break-even prices of two ISPs are:

$$\begin{aligned} p_1^*(\infty) &= \frac{q}{n\alpha_1}(1 - y_2^*) = \frac{q}{n\alpha_1}(\frac{1}{2} + t), \\ p_2^*(\infty) &= \frac{q}{n\alpha_2}(1 - y_1^*) = \frac{q}{n\alpha_2}(\frac{1}{2} - t). \end{aligned} \tag{4.13}$$

Based on the above routing solution, we can answer the question about the profitability and survivability of the ISPs in the following proposition.

Proposition 3 *In order for ISP_1, the ISP with larger market share, to benefit more than ISP_2 from peering, the necessary condition in Equation 4.14 must be satisfied.*

$$\alpha_1(c_1^o - c_1^p) < \alpha_2(c_2^o - c_2^p) - 2(\alpha_1 - \alpha_2). \tag{4.14}$$

In order for ISP_1 to maintain a lower break-even price after peering, the necessary condition in Equation 4.15 must be satisfied.

$$\alpha_1(c_1^o - c_1^p) < \alpha_2(c_2^o - c_2^p) + 2(\alpha_1 - \alpha_2). \tag{4.15}$$

Proof:

From Equation 4.13, we know the break-even prices after unlimited capacity peering are:

$$p_1^*(\infty) = \frac{q}{n\alpha_1}(\frac{1}{2} + t), \qquad p_2^*(\infty) = \frac{q}{n\alpha_2}(\frac{1}{2} - t).$$

This allows us to write down the benefits for the ISPs as:

$$\begin{aligned} p_1^*(0) - p_1^*(\infty) &= \frac{q}{n\alpha_1}(1 - (x_1 + y_1)) = \frac{q}{n\alpha_1}(\frac{1}{2} - t), \\ p_2^*(0) - p_2^*(\infty) &= \frac{q}{n\alpha_2}(1 - (x_2 + y_2)) = \frac{q}{n\alpha_2}(\frac{1}{2} + t). \end{aligned}$$

To let ISP_1 benefit more than ISP_2, we must have:

$$\begin{aligned} &\frac{q}{n\alpha_1}(\frac{1}{2} - t) > \frac{q}{n\alpha_2}(\frac{1}{2} + t) \quad \Rightarrow \quad t < \frac{\alpha_2 - \alpha_1}{2} \\ \Rightarrow \;& \alpha_1(c_1^o - c_1^p) < \alpha_2(c_2^o - c_2^p) - 2(\alpha_1 - \alpha_2). \end{aligned}$$

In order for ISP_1 to have a lower break-even price, we must have:

$$\begin{aligned} &\frac{q}{n\alpha_1}(\frac{1}{2} + t) < \frac{q}{n\alpha_2}(\frac{1}{2} - t) \quad \Rightarrow \quad t < \frac{\alpha_1 - \alpha_2}{2} \\ \Rightarrow \;& \alpha_1(c_1^o - c_1^p) < \alpha_2(c_2^o - c_2^p) + 2(\alpha_1 - \alpha_2). \end{aligned}$$

∎

Both Equation 4.14 and Equation 4.15 are expressed in terms of $(c_i^o - c_i^p)$, the difference between the capacities of two incoming links for a given ISP. Both conditions require a certain parity of incoming link capacity differences between different competing ISPs.

Before peering, ISP_1 has a lower break-even price than that of ISP_2, so that sometimes ISP_1 may still maintain a lower break-even price after peering even if ISP_2 achieves more benefit from peering than ISP_1. In this case, peering might still be a reasonable option for ISP_1 since it does not cause ISP_1 to lose its break-even pricing advantage over ISP_2. As expected, the condition of Equation 4.14 to extract more benefit from peering is more stringent than the condition of Equation 4.15 to maintain price advantage, since $2(\alpha_1 - \alpha_2) > 0$.

We enumerate the different possible outcomes after unlimited capacity peering as follows.

1. ISP_1 maintains a lower break-even price than ISP_2, and achieves more benefits than ISP_2 from peering,
2. ISP_1 maintains a lower break-even price than ISP_2, but ISP_2 achieves more benefits than ISP_1 from peering,
3. ISP_1 achieves less benefits from peering, and has a higher break-even price than ISP_2.

This order is also the preference for ISP_1. To visualize, let us assume $\alpha_1 = 3/5$ and depict the outcomes under different values of $(c_1^o - c_1^p)$ and $(c_2^o - c_2^p)$ in Figure 4.5.

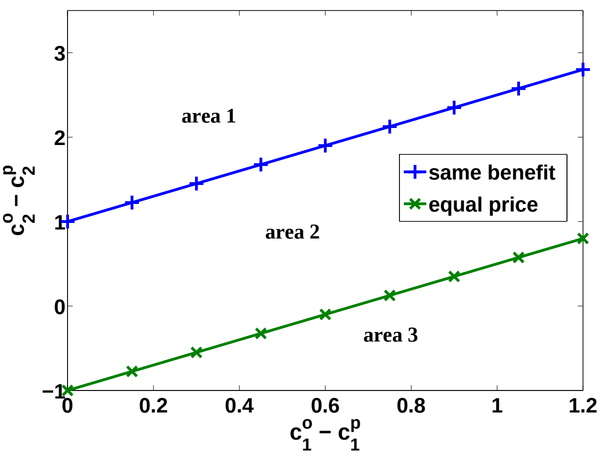

Figure 4.5: Different Outcomes of Unlimited Capacity Peering.

From Equation 4.13 we see that there are at least two issues with unlimited capacity peering:

- It is in each ISP's selfish interest to lower the provisioning of its private link and uploading capacity of the peering link, in order to benefit more from the unlimited capacity peering. For a given c_2^o and c_1^p, ISP_1 can successively lower c_1^o or c_2^p to achieve better and better outcome for itself. From an overall network point of view, this tends to discourage network growth and service improvement. After a point, the subscriber demand cannot be satisfied.

- Each ISP's destiny may be affected by the other ISP's provisioning. For example, assuming c_1^o is fixed by ISP_1 and two ISPs have negotiated the capacity of the peering link. When ISP_2 lowers c_2^o, the outcome for ISP_1 may still drop from area 1 to 2, to 3. The reason for this lost of control is because application layer routed traffic under unlimited peering makes the two ISP networks behave like a single network. Note that before application layer routing appeared, we did not have this problem with web traffic. With only web/email traffic, ISPs have more control over traffic routing through policy-based routing.

Limiting the peering link capacity c_1^p and c_2^p might be one of the remedies available to ISPs. When the capacity of the peering link is sufficiently reduced, we have the routing solution $y_1^* + y_2^* < 1$, thus limiting the usage of an ISP's private link by other ISPs. This scenario is studied in the next subsection.

4.4.3 Limited Capacity Peering Scenario

In this case, the optimization problem in Equation 4.7 can be transformed to the following format:

$$\begin{aligned}\mathbf{min}_{y_1,y_2} \quad & \alpha_2\left(y_1 - \frac{c_2^p - c_2^o + 2}{4}\right)^2 \\ + \quad & \alpha_1\left(y_2 - \frac{c_1^p - c_1^o + 2}{4}\right)^2, \\ \textbf{subject to} \quad & l_i \le y_i \le h_i.\end{aligned} \tag{4.16}$$

The optimal solution is

$$y_1^* = \begin{cases} y_1^{'} & : \text{ if } y_1^{'} \in [l_1, h_1] \\ l_1 & : \text{ if } y_1^{'} < l_1 \\ h_1 & : \text{ if } y_1^{'} > h_1. \end{cases} \tag{4.17}$$

We can solve for y_2^* in the same way as y_1^*. And the break-even prices of two ISPs are:

$$\begin{aligned} p_1^* &= \frac{q}{n\alpha_1}(1 - y_2^*) = \frac{q}{n\alpha_1}\left(\frac{2 - c_1^p + c_1^o}{4}\right), \\ p_2^* &= \frac{q}{n\alpha_2}(1 - y_1^*) = \frac{q}{n\alpha_2}\left(\frac{2 - c_2^p + c_2^o}{4}\right). \end{aligned} \tag{4.18}$$

Similar to the unlimited capacity peering scenario, the following proposition gives the ISPs' economic predicaments in the limited capacity peering scenario:

Proposition 4 *In order for ISP_1, the ISP with larger market share, to benefit more than ISP_2 from limited capacity peering, the necessary condition in Equation 4.19 must be satisfied.*

$$\alpha_2(c_1^o - c_1^p) < \alpha_1(c_2^o - c_2^p) - 2(\alpha_1 - \alpha_2). \tag{4.19}$$

In order for ISP_1 to maintain a lower break-even price after peering with limited capacity, the condition in Equation 4.20 must be satisfied.

$$\alpha_2(c_1^o - c_1^p) < \alpha_1(c_2^o - c_2^p) + 2(\alpha_1 - \alpha_2). \tag{4.20}$$

Proof:

From Equation 4.18, we know the break-even prices after limited capacity peering are

$$p_1^* = \frac{q}{n\alpha_1}(\frac{2 - c_1^p + c_1^o}{4}), \quad p_2^* = \frac{q}{n\alpha_2}(\frac{2 - c_2^p + c_2^o}{4}).$$

This allows us to write down the benefits for ISPs as:

$$p_1^*(0) - p_1^* = \frac{q}{n\alpha_1}(1 - (x_1 + y_1)) = \frac{q}{n\alpha_1}\frac{c_1^p - c_1^o + 2}{4},$$
$$p_2^*(0) - p_2^* = \frac{q}{n\alpha_2}(1 - (x_2 + y_2)) = \frac{q}{n\alpha_2}\frac{c_2^p - c_2^o + 2}{4}.$$

To let ISP_1 benefit more than ISP_2, we must have:

$$\frac{q}{n\alpha_1}\frac{c_1^p - c_1^o + 2}{4} > \frac{q}{n\alpha_2}\frac{c_2^p - c_2^o + 2}{4}$$
$$\Rightarrow \alpha_2(c_1^o - c_1^p) < \alpha_1(c_2^o - c_2^p) - 2(\alpha_1 - \alpha_2).$$

In order for ISP_1 to have a lower break-even price, we must have:

$$\frac{q}{n\alpha_1}\frac{2 - c_1^p + c_1^o}{4} < \frac{q}{n\alpha_2}\frac{2 - c_2^p + c_2^o}{4}$$
$$\Rightarrow \alpha_2(c_1^o - c_1^p) < \alpha_1(c_2^o - c_2^p) + 2(\alpha_1 - \alpha_2).$$

∎

Again, in order for ISP_1 to maintain certain economic advantages, it still has the incentive to lower its provisioning and it must try to maintain certain parity of the difference in incoming link capacities with that of its competitor. However, in this case the conditions are further relaxed. For example, condition in Equation 4.19 is a more relaxed version of the condition in Equation 4.14 since $\alpha_1 \geq \alpha_2$. This reveals the fact that it is easier for ISP_1 to maintain economic advantages under limited capacity peering scenario than unlimited capacity peering scenario. It also explains part of the reason why large ISPs want to limit the capacity of their peering links.

Furthermore, from Equation 4.18, we can see that with limited peering, each ISP's break-even price can be determined by its own decision and would not be affected by the other ISP's private link provisioning. This advantage is also likely to make one ISP to favor limited capacity peering to gain more control of its own fate.

4.5 Numerical Example

In this section, we illustrate and explain the results in Section 4.4 through a simple numerical example. We assume $\alpha_1 = \frac{3}{5}$, $\alpha_2 = \frac{2}{5}$, $c_2^o = 0.9$, and $c_1^p = 1.5$. All these parameter are fixed. We will study ISP_1's provisioning in this section.

4.5.1 Private Link Provisioning

Let us further assume $c_2^p = c_1^p = 1.5$. We will focus on ISP_1's private link provisioning and study how c_1^o affects ISP_1's break-even price. The routing problem under these parameters is:

$$\begin{aligned} \mathbf{min}_{y_1,y_2} \quad & 0.4(y_1 - 0.65)^2 + 0.6(y_2 - \frac{3.5 - c_1^o}{4})^2, \\ \textbf{subject to} \quad & 0.1 \le y_1 \le 1, \\ & \max(0, 1 - c_1^o) \le y_2 \le 1, \\ & y_1 + y_2 \le 1. \end{aligned} \tag{4.21}$$

Given $c_2^o = 0.9$, we must have $c_1^o \geq 0.1$ in order to satisfy the total traffic demand. We solve this problem and plot p_1^* and p_2^* in Figure 4.6 for $c_1^o \geq 0.1$. It shows ISP_1's beak-even price

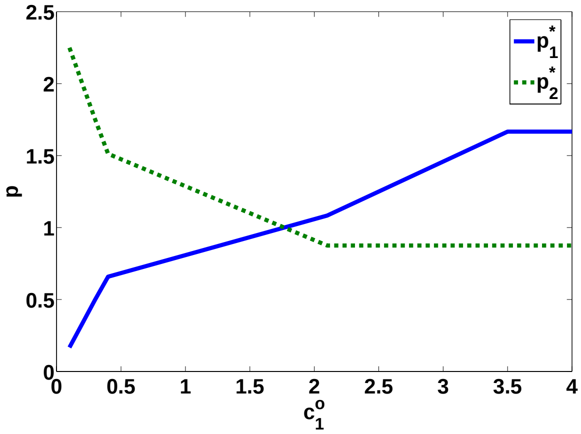

Figure 4.6: Break-even Price of Two ISPs as We Vary c_1^o.

p_1^* continues to increase as c_1^o increases. This is expected since ISP_1's operating cost increases. However, ISP_2's break-even price p_2^* decreases as c_1^o increases since ISP_2 also takes advantages of ISP_1's increased private link capacity without any cost, due to the routing of the overlay network. In real world, this phenomenon may not be that apparent because overlay streaming is not the only traffic. However, with the increasing popularity and intensity of overlay traffic, it would have more influence on ISPs' operation.

4.5.2 Peering Link Provisioning

Now let us fix $c_1^o = 1.1$ and study the influence of different c_2^p on two ISPs' break-even prices. The routing problem under these parameters is:

$$\begin{aligned}
\mathbf{min}_{y_1,y_2} \quad & 0.4(y_1 - \frac{c_2^p + 1.1}{4})^2 + 0.6(y_2 - 0.6)^2, \\
\textbf{subject to} \quad & 0 \le y_2 \le 1, \\
& 0.1 \le y_1 \le \min(1, c_2^p), \qquad (4.22) \\
& y_1 + y_2 \le 1.
\end{aligned}$$

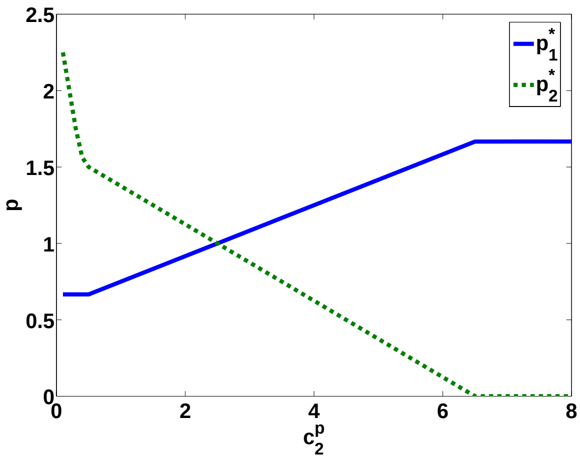

Figure 4.7: Break-even Price of Two ISPs as We Vary c_2^p.

After solve this optimization problem, we plot p_1^* and p_2^* in Figure 4.7. From it we can see ISP_1's break-even price increases while ISP_2's break-even price decreases as ISP_1 increases the capacity of the peering link from ISP_1 to ISP_2. Because of the "free-ride" phenomenon, ISP_2 can take more advantages from the peering link when c_2^p increases.

This example shows that an ISP must be very careful when provisioning its private links and peering links. Otherwise, other ISPs will be able to *free-ride* and gain an economic advantage over it.

4.6 Markets with Multiple ISPs

Now let us generalize the above analysis to a market with n ISPs. Similar as before, the market share of ISP_i is defined as α_i. c_i^o is the capacity of the provider link from the Internet to ISP_i, which is referred to as ISP_i's private capacity. We assume all these ISPs are free peering with each other, while the details of peering capacities are ignored. Instead, we simply use c_i^p to denote the total capacity of all the peering links from other ISPs to ISP_i.

The routing of the overlay demand is characterized by x_i, which is the fraction of the demand downloaded by ISP_i from the Internet, and $y_i = 1 - x_i$, which is the fraction that ISP_i downloads from other ISPs through its peering links.

The performance of the local network can be measured by the following objective function:

$$\begin{aligned}
&\sum_{i=1}^{n}(\alpha_i(c_i^o - x_i)x_i + \alpha_i(c_i^p - y_i)y_i) \\
= &\sum_{i=1}^{n}\alpha_i(-2(x_i^2 - \frac{c_i^o - c_i^p + 2}{2}x_i) + c_i^p - 1). \qquad (4.23)
\end{aligned}$$

The capacity and demand constraints are as follows:

$$x_i \leq c_i^o, \qquad y_i \leq c_i^p. \qquad (4.24)$$

4.6.1 Unlimited Capacity Peering Scenario

If the capacities of peering links in the local market are large enough, so that all demand is downloaded by local ISPs only once and then shared among them, we say it is "unlimited peering scenario", *i.e.* $\sum_{i=1}^{n} x_i = 1$. The optimization problem for the routing in this scenario is as follows.

$$\begin{aligned}
min_{x_i} \quad &\sum_{i=1}^{n}\alpha_i(x_i^2 - \frac{c_i^o - c_i^p + 2}{2}x_i) \\
\text{subject to} \quad &\sum_{i=1}^{n} x_i = 1.
\end{aligned}$$

The solution of this problem is $x_i = (\frac{\kappa}{\alpha_i} + \frac{c_i^o - c_i^p + 2}{4})$, where $\kappa = \frac{\sum(c_i^p - c_i^o) + 4 - 2n}{4\sum 1/\alpha_i}$. As a result, the break-even price for

ISP_i is

$$p_i(\infty) = \frac{q}{n\alpha_i} x_i = \frac{q}{n\alpha_i}(\frac{\kappa}{\alpha_i} + \frac{c_i^o - c_i^p + 2}{4}). \quad (4.25)$$

The benefit ISP_i achieves from the free peering is

$$p_i(0) - p_i(\infty) = \frac{q}{n\alpha_i}(1 - x_i) = \frac{q}{n\alpha_i}(\frac{c_i^p - c_i^o + 2}{4} - \frac{\kappa}{\alpha_i}). \quad (4.26)$$

Therefore, the condition for that ISP_i has a lower break-even price than ISP_j is

$$\frac{1}{\alpha_i}(\frac{\kappa}{\alpha_i} + \frac{c_i^o - c_i^p + 2}{4}) < \frac{1}{\alpha_j}(\frac{\kappa}{\alpha_j} + \frac{c_j^o - c_j^p + 2}{4})$$
$$\Rightarrow \alpha_j(c_i^o - c_i^p) < \alpha_i(c_j^o - c_j^p) + (\alpha_i - \alpha_j)((\frac{1}{\alpha_i} + \frac{1}{\alpha_j})4\kappa + 2).$$

And the condition for that ISP_i achieves more benefit from the peering than ISP_j is

$$\frac{q}{n\alpha_i}(\frac{c_i^p - c_i^o + 2}{4} - \frac{\kappa}{\alpha_i}) > \frac{q}{n\alpha_j}(\frac{c_j^p - c_j^o + 2}{4} - \frac{\kappa}{\alpha_j})$$
$$\Rightarrow \alpha_j(c_i^o - c_i^p) < \alpha_i(c_j^o - c_j^p) + (\alpha_i - \alpha_j)((\frac{1}{\alpha_i} + \frac{1}{\alpha_j})4\kappa - 2).$$

As a summary, we have the following proposition:

Proposition 5 *In order for ISP_i to benefit more than ISP_j from unlimited capacity peering, the following condition must be satisfied.*

$$\alpha_j(c_i^o - c_i^p) < \alpha_i(c_j^o - c_j^p) + (\alpha_i - \alpha_j)((\frac{1}{\alpha_i} + \frac{1}{\alpha_j})4\kappa - 2). \quad (4.27)$$

In order for ISP_i to maintain a lower break-even price after peering with unlimited capacity, the following condition must be satisfied.

$$\alpha_j(c_i^o - c_i^p) < \alpha_i(c_j^o - c_j^p) + (\alpha_i - \alpha_j)((\frac{1}{\alpha_i} + \frac{1}{\alpha_j})4\kappa + 2). \quad (4.28)$$

When $n = 2$, $i = 1$ and $j = 2$, the above two conditions are consistent with Equation 4.19 and Equation 4.20.

4.6.2 Limited Capacity Peering Scenario

In this scenario, we do not have the constraint $\sum_{i=1}^{n} x_i = 1$, and x_i can be viewed as independent variables. We can solve the routing problem easily and get the solution $x_i = \frac{c_i^o - c_i^p + 2}{4}$.

Therefore, the break-even price for ISP_i is

$$p_i = \frac{q}{n\alpha_i} x_i = \frac{q}{n\alpha_i} \frac{c_i^o - c_i^p + 2}{4}. \tag{4.29}$$

The benefit ISP_i achieves from the free peering links is

$$\frac{q}{n\alpha_i}(1 - x_i) = \frac{q}{n\alpha_i} \frac{c_i^p - c_i^o + 2}{4}. \tag{4.30}$$

As a result, the condition for that ISP_i has a lower break-even price than ISP_j is:

$$\begin{aligned} p_i < p_j \quad &\Rightarrow \quad \frac{1}{\alpha_i} \frac{c_i^o - c_i^p + 2}{4} < \frac{1}{\alpha_j} \frac{c_j^o - c_j^p + 2}{4} \\ &\Rightarrow \quad \alpha_j (c_i^o - c_i^p) < \alpha_i (c_j^o - c_j^p) + 2(\alpha_i - \alpha_j) \end{aligned}$$

And the condition for that ISP_i achieves more benefit from the peering than ISP_j is

$$\begin{aligned} &\frac{q}{n\alpha_i} \frac{c_i^p - c_i^o + 2}{4} > \frac{q}{n\alpha_j} \frac{c_j^p - c_j^o + 2}{4} \\ \Rightarrow \quad &\alpha_j (c_i^o - c_i^p) < \alpha_i (c_j^o - c_j^p) - 2(\alpha_i - \alpha_j). \end{aligned}$$

As a summary, we have the following proposition:

Proposition 6 *In order for ISP_i to benefit more than ISP_j from limited capacity peering, the following necessary condition must be satisfied.*

$$\alpha_j (c_i^o - c_i^p) < \alpha_i (c_j^o - c_j^p) - 2(\alpha_i - \alpha_j). \tag{4.31}$$

In order for ISP_i to maintain a lower break-even price after peering with limited capacity, the following necessary condition must be satisfied.

$$\alpha_j(c_i^o - c_i^p) < \alpha_i(c_j^o - c_j^p) + 2(\alpha_i - \alpha_j). \tag{4.32}$$

This proposition is a generalized version of Proposition 4.

4.7 Summary

The emergence of application layer routing technologies such as P2P overlays is a relatively abrupt and new phenomenon. Since there is a growing trend that users are adopting better technologies to gain broadband access, overlay demand such as P2P traffic will be more prominent. The implications of this phenomenon on the traffic engineering and the Internet operation have not been investigated before. Therefore, the analysis in this chapter not only provides the fundamental understanding on why there might be a shift on the Internet operation, but also open doors for potential research on overlay traffic management, peering relationship establishment policy, as well as routing decision among autonomous systems.

In this chapter we conduct an economic analysis for an overlay streaming network to gain some insights on the free ride phenomenon brought by overlay networks. To capture the important difference between how overlay traffic and traditional web traffic are routed in the Internet, we model the routing in the network as to optimize the performance of transferring each byte of its subscribers' demand. We deploy break-even price as the metric to evaluate the economic efficiency of one ISP. The analysis shows that ISPs must control their provider link capacities as well as free-peering link capacities they provide to their peering ISPs to maintain certain parity between different incoming link capacities among all ISPs. Only when such parity is maintained, ISPs can sustain their relative economic positions despite of application layer routing. It reveals the fact that the application layer routing of overlay networks upsets ISPs' traditional peering strategy.

Particularly, we (a) reveal the phenomenon that interdomain routing protocol cannot control the traffic routed at application layer, (b) show that peering relationship with unlimited capacity (or very high capacity) discourages network growth and service improvement if overlay traffic dominates, (c) show that while peering with limited capacity may reduce ISPs' benefit from economy of scale, but it can remove or minimize the "fairness" problem in distributing the peering benefits.

In fact, there are many other kinds of overlay traffic that may influence the traffic engineering policy and peering decision. For example, in order to optimize performance or save money, some people set up "proxies" to redirect their interdomain traffic to go through one ISP that is different from the default BGP routing. Such traffic is also a kind of "application layer routed" traffic. Another example is routing overlay network. In the next chapter, we will provide a general framework for all kinds of traffic and then conduct an economic analysis for the peering strategy and provisioning strategy of ISPs in a dynamic market.

□ **End of chapter.**

Chapter 5

Inter-ISP Traffic and Peering Strategy

Summary

Inter-ISP traffic flow determines the settlement between ISPs and affects the perceived performance of ISPs' services, therefore it has an important influence on ISPs' business. In this chapter, we present a game theoretic framework to analyze the economic implications of various inter-ISP traffic patterns on ISPs' peering strategy. Our game theoretic framework captures the decision of the major players of the Internet market - the subscribers and the ISPs playing different transit roles. The analysis is based on a general traffic model that takes into consideration both the network layer routing controlled by ISPs and application layer routing controlled by applications. In particular, we apply this framework to a local market where two ISPs compete for market share of subscribers under two traffic patterns: "web" and "P2P overlay", that typifies the transition the current Internet is going through. Our numerical analysis quantitatively demonstrates some important observations.

Previously, BGP ensures that the actual information flows (traffic) are confined to be along the routes that are advertised between the peering ISPs according to ISPs' business considerations. So ISPs' settlement is what they expect from their negotiated peering agreements. However, policy-based routing focuses on business considerations and does not imply the best possible service for individual subscribers and the applications they are running. As a complement, today's Internet is full of overlay applications, which change traffic routing in the application layer to better satisfy and optimize for applications' (or subscribers') needs.

Inter-ISP traffic exchanges are generated by various applications. Therefore, in today's Internet the inter-ISP traffic flow patterns are controlled not only by ISPs' policy-based routing configuration and traffic engineering, but also by application layer routing of various overlay networks. In Chapter 4, we point out that this phenomenon upsets ISPs' traditional business model and the free peering benefit distribution might be unfair in networks with overlay traffic.

However, that analysis cannot be applied directly on other overlay traffic types since we only focus on one particular overlay traffic model. More importantly, that analysis is conducted in a static market, where both the market share distribution and the provisioning of all links are fixed. As a result, that economic analysis does not take the response of subscribers into consideration, and the evaluation based on break-even price only reveals the influence direction of overlay traffic at a particular point of the market, but cannot capture the stable (final) state of a dynamic market.

The local access market is dynamic in real world. For example, subscribers can shift to other ISPs. An ISP can make peering decisions (*e.g.* whether to peer with the other ISP or not and with what kind of peering relationship), provisioning decisions (*e.g.* how much capacity shall this ISP provision), and pricing decisions (*e.g.* how much shall this ISP charge its own subscribers) to cope with inter-ISP traffic and optimize its business.

In this chapter, we present a general inter-ISP traffic model together with a game-theoretic framework to further understand the ISP peering and provisioning issues under application layer routing. The game-theoretic framework captures the decision of the major players of the Internet marketplace - the subscribers (users) and the ISPs playing different roles in providing transit service, and provides *Nash Equilibrium* as an output to show the steady state of a dynamic market.

We still focus on the situation in a local market where there are two ISPs competing for subscribers and they are connected to the Internet via upstream transit ISPs. In our game theoretic framework, the ISPs are modelled as *leaders* who optimize their profits by setting proper prices, and subscribers are modeled as *followers* who react to the prices set by the local ISPs. The operation of these local ISPs can be expected to reach some market equilibrium, where each ISP attaining a certain market share of the subscribers and making a certain amount of profit.

We adopt a gravity traffic model as a simple and intuitive way to model the inter-ISP traffic patterns generated by different applications. In this model, traffic flows according to the number of subscribers in each ISP engaged in the same application at the same time, modulated by performance influence which leads to specific route selection by overlay networks. In other words, an ISP with more subscribers tends to attract (or contribute) more traffic, with the throttling of the performance of links connecting this ISP to its neighbors. By adjusting its parameters, such a gravity model provides a rational approximation of different inter-ISP traffic patterns. Our inter-ISP traffic model pinpoints the important factors that affect inter-ISP traffic intensities, *i.e.* population distribution, routing sensitivity to network performance, and routing policies which are related to ISPs' business relationship.

Our analysis brings out several interesting observations.

1. As the traffic model shifts from traditional web traffic to more overlay traffic, the distribution of the benefit of the free peering may become unfairly distributed. The smaller ISP can enjoy a certain degree of *free riding*. This may trigger some actions by the bigger ISP, though the possible actions may be counter-productive (as explained below).

2. If the larger ISP tries to cut back the bandwidth provisioning of its private link(s) to the Internet, it will help restore the fairness of peering benefit distribution; but this is detrimental to business expansion and is not healthy for the growth of the Internet.

3. If the larger ISP tries to reduce the peering bandwidth, it will also help restore the fairness of peering benefit distribution. However, this also may hurt the ISP business since it reduces the collective bargaining power of the ISPs in setting prices for subscribers.

5.1 Business Model

We still focus on the same market shown in Figure 4.2. For easy reading, we present this market in Figure 5.1 again. In the previous analysis, our traffic model only includes remote traffic demand from the Internet to the local subscribers. In order to make our analysis complete, in this chapter we will use an abstract traffic model to cover all the demands of subscribers in one ISP: *local demand*, *peering demand* and *remote demand*.

Let D_i denote the total amount of ISP_i's subscribers's demand. Obviously we have $t_i^o + t_i^p < D_i$, which means some of the demand is satisfied by peers or servers in ISP_i's local network. Let t_i^l denote the demand intensity that is satisfied locally, and obviously we have $t_i^l = D_i - t_i^o - t_i^p$.

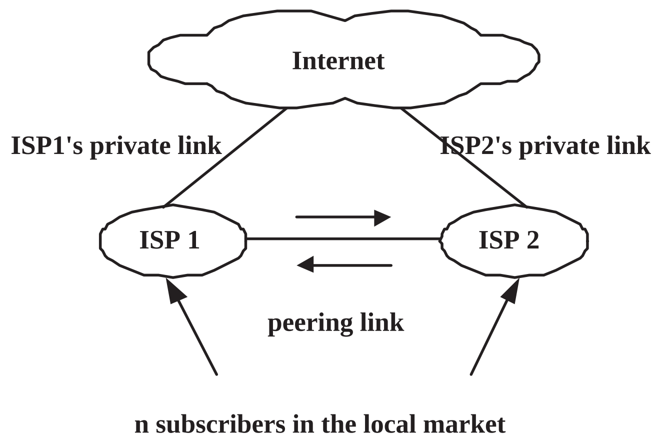

Figure 5.1: The Local Market under Study.

Local ISPs pay their providers for using their transit service. We assume that the transit providers charge local ISPs in the same local market, (*i.e.* ISP_1 and ISP_2), in the same way. The cost incurred is computed based on the amount of traffic a customer ISP generates and the performance of the link between two ISPs, *i.e.* $cost = f(c, t)$, where t is a variable determined by the customer ISP's traffic volume, c is related to the capacity of the inter-ISP links and $f(\cdot)$ is a non-decreasing function that maps t and c to cost.

Various charging models differ from one another in their choices of c, t and the cost function $f(\cdot)$. For example, t can be the total volume of traffic one customer ISP generates during the entire charging period, which is referred to as "total-volume based charging"; or t can be the 95th percentile of all 5-minute traffic volumes during the charging period, which is called "percentile-based charging". In this analysis, we are not going to study the traffic dynamics on a short time scale, which implies that t can be the average traffic intensity on the inter-ISP links for our purpose.

Although c does not usually appear directly in the cost function, it in fact affects the cost which can be seen from the comparison that a traffic intensity of $40Mbps$ on a OC3 link must cost more than a traffic intensity of $40Mbps$ on a DS3 link.

The cost function $f(\cdot)$ can be quite complicated in the real world, for example, a piece-wise linear (non-decreasing) function.

5.2 Game Theoretic Framework

In our game formulation for the network shown in Figure 5.1, local ISPs have determined the capacity of the links in the network, *i.e.* c_i^o and c_i^p are fixed. The game is to model the interaction of two ISPs and subscribers in this market, and find out the market share α_i and the *optimum* price p_i when the game for this market reaches *Nash Equilibrium*. Later we will analyze Nash Equilibria of games with different c_i^o and c_i^p to study ISPs' peering strategies and provisioning strategies.

In the local market, ISPs set the prices for their own subscribers independently, and each subscriber responds to the prices by choosing one ISP as its provider to optimize its own utility. It is naturally a two-stage multi-leader-follower game [20], in which ISPs act as leaders, with n subscribers acting as followers that play a game with the prices $p_i (i = 1, 2)$ as ISP strategies.

Similar to the analysis of a Stackelberg game, we first analyze the second stage of this game, assuming that ISP_i $(i = 1, 2)$ has set the price p_i at the first stage of the game. All subscribers in this game share the same utility function, which means they choose their individual providers according to the same rules.

The most important concerns of one subscriber are access cost (p_i if choose ISP_i) and service performance. Since one ISP is with fixed inter-ISP link capacity, its service performance will become worse as it provides access service to more subscribers. Let $P_i(\alpha_i)$ denote the service performance of ISP_i when it provides access service to $n\alpha_i$ subscribers. Obviously $P_i(\cdot)$ is a decreasing function and $P_1(\cdot)$ may not be the same as $P_2(\cdot)$ because two ISPs may have different inter-ISP link capacity.

Let $\mathcal{U}_i(p_i, \alpha_i)$ denote the utility value of subscribers in ISP_i when ISP_i provides service for $n\alpha_i$ subscribers at a price of p_i. We define

$$\mathcal{U}_i(p_i, \alpha_i) = P_i(\alpha_i) - p_i, \tag{5.1}$$

which shows that subscribers always prefer lower access cost and better service performance. Since all subscribers of one ISP are with the same utility value, we also refer to $\mathcal{U}_i$ as *ISP_i's service utility*.

Subscribers are not required to make their decisions simultaneously, which implies that we consider the Nash Equilibrium of the second stage game as a stable state where all subscribers settle down and will not change their decisions. Let α_1^* be ISP_1's market share at the Nash Equilibrium point. After all subscribers in the local market finish selecting their ISPs, we have

$$\mathcal{U}_1(p_1, \alpha_1^*) = \mathcal{U}_2(p_2, 1 - \alpha_1^*), \tag{5.2}$$

which means the utility of all subscribers in both ISPs evaluate to the same value, hence none of them has a reason to change his decision and thus the market becomes stable.

Note that Equation 5.2 may not hold for some extreme cases with $\alpha_1^* = 0$ or $\alpha_1^* = 1$. For example, $\mathcal{U}_1(p_1, 1) > \mathcal{U}_2(p_2, 0)$ implies ISP_1 grabs all subscribers and it still can provide a better utility for new subscribers. This is a degenerate case where one ISP dies. In the following analysis, we ignore such degenerate cases, and assume Equation 5.2 always holds for Nash Equilibrium of the game.

Let us define $\varphi(\alpha) = P_1(\alpha) - P_2(1-\alpha)$. From Equation 5.1 and Equation 5.2, we have

$$p_1 - p_2 = P_1(\alpha_1^*) - P_2(1-\alpha_1^*) = \varphi(\alpha_1^*).$$

We assume that the two local ISPs together have the full control of the market by setting the prices, which is equivalent to assuming that $\varphi(\cdot)$ is an invertible function. Then we can solve the second stage of our game and the Nash Equilibrium is

$$\alpha_1^*(p_1, p_2) = \varphi^{-1}(p_1 - p_2). \tag{5.3}$$

In summary, the second stage game takes prices set by ISPs (*i.e.* p_1 and p_2) as input and gives $\alpha_1^*(p_1, p_2)$ as outcome. $\alpha_i^*(p_1, p_2)$ reflects the rational response of subscribers to the prices p_1 and p_2.

At the first stage of the game, the ISPs do not cooperate among themselves and they are competing with each other for market share. They need to set their prices (in a one-shot game) simultaneously and independently, which implies the first stage game is a standard Nash game. The action set of ISP_i in this non-cooperate game is $p_i \geq 0$. We define a utility function $\mathcal{R}_i$, where $\mathcal{R}_i(p_1, p_2)$ is the preference of ISP_i to an action profile (p_1, p_2). $\mathcal{R}_i(p_1^{'}, p_2^{'}) > \mathcal{R}_i(p_1^{''}, p_2^{''})$ means ISP_i prefers the profile $(p_1^{'}, p_2^{'})$ to $(p_1^{''}, p_2^{''})$.

We assume that one ISP simply uses its profit (revenue minus cost) to evaluate a profile. ISPs, acting as leaders of the multi-leader-follower game, are aware of the function $\alpha_i^*(p_1, p_2)$, the rational reaction of subscribers to prices. Therefore the utility function is defined as follows:

$$\mathcal{R}_i(p_1, p_2) = p_i n \alpha_i^*(p_1, p_2) - f(t_i^o(\alpha_i^*(p_1, p_2)), c_i^o). \tag{5.4}$$

Here, t_i^o, the traffic intensity on ISP_i's private link, is certainly related to ISP_i's subscriber share, so we write it in the form of $t_i^o(\alpha_i^*(p_1, p_2))$.

From Equation 5.3, we can see that the market share distribution would not change if two ISPs agree to increase their prices at the same time, and then both ISPs can increase their revenues. However, two ISPs are not cooperating with each other, and both of them have incentives to lower their individual price to attract subscribers. In the numerical studies, we will show that peering helps two ISPs cooperate with each other to some extent and it results in higher prices.

As a summary, the multi-leader-follower game played by two ISPs and subscribers gives rise to the following optimization problem:

$$\begin{array}{lll} \max_{p_i \geq 0} & \mathcal{R}_i(\hat{p_i}), & i = 1, 2 \\ \text{subject to} & \alpha_1^*(p_1, p_2) = \varphi^{-1}(p_1 - p_2), & \end{array} \tag{5.5}$$

where $\hat{p_1} = (p_1, p_2^*)$ and $\hat{p_2} = (p_1^*, p_2)$.

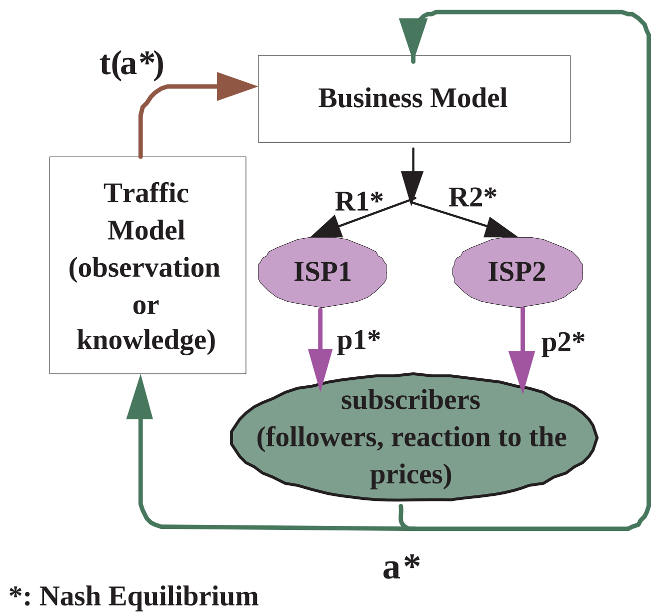

Figure 5.2: Overview of the Game Theoretic Framework.

In Figure 5.2, we give an overview of the whole game theoretic framework to study the local market. For any given p_1 and p_2, subscribers play a game as n followers and give α_i as an outcome, which is reflected in the constraint in Equation 5.5. ISP_1 and ISP_2 can predict the response α_i under different profiles (p_1, p_2), and then they can predict their own profits ($\mathcal{R}_1$ and $\mathcal{R}_2$) under different profiles according to the business model. Therefore, two ISPs that are playing a non-cooperate game are able to maximize their own profits by making optimum decisions on their own prices. The interaction of two ISPs gives (p_1^*, p_2^*) as the Nash Equilibrium which represents the steady state of the market.

5.3 Solution of the Game

As an extension of the Nash game which models the situation where each player takes no leadership position over its rivals, the multi-leader-follower game arises to analyze the situation where one or more players have more power than other players and thus become leader(s) in the game. The simplest of multi-leader-follower game is the Stackelberg game in which there is only one leader and multiple followers. A mathematical model for the Stackelberg game is the MPEC (mathematical programs with equilibrium constraints). The computation of global solutions to MPECs remains elusive, if not impossible, regardless of the progress in research on the MPEC [17] [29].

A multi-leader-follower game with multiple leaders is more complicated than a Stackelberg game. The existence of Nash Equilibrium for a multi-leader-follower game is in jeopardy. Even in the favorable case where such an equilibrium exists, its complete characterization remains a daunting, if not impossible, task. Any rigorous attempt to compute a Nash Equilibrium (if it exists) for a multi-leader-follower game is presently out of the reach of existing methods, due to the high level of complexity and technical difficulties [20] [42].

Fortunately, the game in this paper takes specific utility functions $\mathcal{U}_i$ and $\mathcal{R}_i$, so that we are able to do a further study on its Nash Equilibrium. First, we rewrite Equation 5.5 as follows:

$$\begin{cases} p_1^* = \text{argmax}_{p_1} & p_1 n\alpha_1(p_1, p_2^*) - f(t_1^o(\alpha_1(p_1, p_2^*)), c_1^o), \\ p_2^* = \text{argmax}_{p_2} & p_2 n\alpha_2(p_1^*, p_2) - f(t_2^o(\alpha_2(p_1^*, p_2)), c_2^o), \\ \text{subject to} & p_1 - p_2 = \varphi(\alpha_1). \end{cases} \quad (5.6)$$

We have the following theorem on the necessary condition for a Nash Equilibrium of the game defined in the above equation:

Theorem 1 *Define*

$$\begin{aligned} \Phi_1(\alpha_1) &= \varphi(\alpha_1) + \alpha_1 \frac{\partial \varphi}{\partial \alpha_1} - \frac{1}{n}\frac{\partial f(c_1)}{\partial t_1^o}\frac{\partial t_1^o}{\partial \alpha_1}, \\ \Phi_2(\alpha_1) &= \varphi(\alpha_1) + (\alpha_1 - 1)\frac{\partial \varphi}{\partial \alpha_1} - \frac{1}{n}\frac{\partial f(c_2)}{\partial t_2^o}\frac{\partial t_2^o}{\partial \alpha_1}. \end{aligned} \quad (5.7)$$

For α_1^* to be a Nash Equilibrium of the multi-leader-follower game defined in Equation 5.6, it must satisfy the following two conditions:

$$\begin{aligned} i). \quad & \Phi_1(\alpha_1^*) + \Phi_2(\alpha_1^*) - \varphi(\alpha_1^*) = 0, \\ ii). \quad & \frac{\partial \Phi_1}{\partial \alpha_1}|_{\alpha_1^*} < 0, \qquad \frac{\partial \Phi_2}{\partial \alpha_1}|_{\alpha_1^*} < 0. \end{aligned}$$

Proof: Assume α_1^* is a Nash Equilibrium that is yielded from (p_1^*, p_2^*). Obviously we have $p_1^* - p_2^* = \varphi(\alpha_1^*)$.

From the definition of Nash Equilibrium, we know that p_1^* should be the best response to p_2^*, and vice versa. Since we do not consider the cases where the maximum profit happens at the boundary, we must have

$$\begin{aligned} \frac{\partial \mathcal{R}_1(p_1, p_2^*)}{\partial p_1}|_{p_1^*} = 0, \qquad \frac{\partial R_1^2(p_1, p_2^*)}{\partial p_1^2}|_{p_1^*} < 0, \\ \frac{\partial \mathcal{R}_2(p_1^*, p_2)}{\partial p_2}|_{p_2^*} = 0, \qquad \frac{\partial R_2^2(p_1^*, p_2)}{\partial p_2^2}|_{p_2^*} < 0. \end{aligned}$$

Moreover, because $\alpha_1 = \varphi^{-1}(p_1, p_2)$ exists, we can rewrite the conditions as follows:

$$\frac{\partial \mathcal{R}_1(\alpha_1, p_2^*)}{\partial \alpha_1}|_{\alpha_1^*} = 0, \qquad \frac{\partial \mathcal{R}_2(\alpha_1, p_1^*)}{\partial \alpha_1}|_{\alpha_1^*} = 0, \tag{5.8}$$

$$\frac{\partial \mathcal{R}_1^2(\alpha_1, p_2^*)}{\partial \alpha_1^2}|_{\alpha_1^*} < 0, \qquad \frac{\partial \mathcal{R}_2^2(\alpha_1, p_1^*)}{\partial \alpha_1^2}|_{\alpha_1^*} < 0. \tag{5.9}$$

From Equation 5.8 and $p_1^* - p_2^* = \varphi(\alpha_1^*)$, we derive that

$$p_2^* + \varphi(\alpha_1^*) + \alpha_1^* \frac{\partial \varphi}{\partial \alpha_1}|_{\alpha_1^*} - \frac{1}{n}\frac{\partial f(c_1^o)}{\partial t_1^o}\frac{\partial t_1^o}{\partial \alpha_1}|_{\alpha_1^*} = 0,$$

$$-p_2^* + (\alpha_1^* - 1)\frac{\partial \varphi}{\partial \alpha_1}|_{\alpha_1^*} - \frac{1}{n}\frac{\partial f(c_2^o)}{\partial t_2^o}\frac{\partial t_2^o}{\partial \alpha_1}|_{\alpha_1^*} = 0.$$

It can be further simplified based on Equation 5.7 as follows:

$$\begin{aligned} p_2^* + \Phi_1(\alpha_1^*) &= 0, \\ -\varphi(\alpha_1^*) - p_2^* + \Phi_2(\alpha_1^*) &= 0. \end{aligned} \tag{5.10}$$

α_1^* must satisfy the following condition in order to be the market share of a Nash Equilibrium:

$$\Phi_1(\alpha_1^*) + \Phi_2(\alpha_1^*) - \varphi(\alpha_1^*) = 0. \tag{5.11}$$

Equation 5.9 can be simplified as

$$\frac{\partial \Phi_1}{\partial \alpha_1}|_{\alpha_1^*} < 0, \qquad \frac{\partial \Phi_2}{\partial \alpha_1}|_{\alpha_1^*} < 0.$$

∎

The theorem gives us a way to compute the Nash Equilibrium of the game which represents the steady state of the local market. The α_1^* can be solved using Equation 5.11, and p_1^* and p_2^* can be derived from $p_2^* = -\Phi_1(\alpha_1^*)$ and $p_1^* = \Phi_2(\alpha_1^*)$.

Equation 5.11 shows that the traffic pattern $t_i^o(\alpha_i)$ affects the Nash Equilibrium. In the next section, we will present our traffic model as a simple and intuitive way to represent inter-ISP traffic patterns. After that we will show how our framework can help ISPs make peering and provisioning decisions in the context of different traffic patterns.

5.4 Inter-ISP Traffic Model

The inter-ISP traffic is a significant factor that determines the outcome of the ISP peering game, for at least two reasons. First, the monetary flows among ISPs are directly based on inter-ISP traffic flows. Secondly, inter-ISP links are likely to be bottlenecks that affect the performance perceived by potential subscribers, hence indirectly affect the ISPs' market share distribution. Therefore t_i^o appears in both subscribers' utility function $\mathcal{U}_i$ and ISPs' utility function $\mathcal{R}_i$. On the other hand, it is very difficult to characterize real-world traffic patterns using simple traffic models. This is usually the most contentious part of modeling efforts.

In this paper, we describe a reasonably simple yet general traffic model $t(\alpha)$, focusing on our problem at hand, namely describing the internal and external traffic of a set of local ISPs based on their market share distribution α and their network provisioning.

Figure 5.3 is an illustration of all factors that affect the intensity of inter-ISP traffic. In a network, ISPs decide the provisioning of their links (c_i^o and c_i^p), and they also implement their routing rules using BGP. The market share distribution (α) is determined by subscribers' choice of their own providers, and subscribers may also affect the routing of the traffic using application layer routing. The traffic model takes all these factors as input and predicts traffic intensities on inter-ISP links as output. The traffic model shows the reaction of the network to the behavior of the ISPs and subscribers.

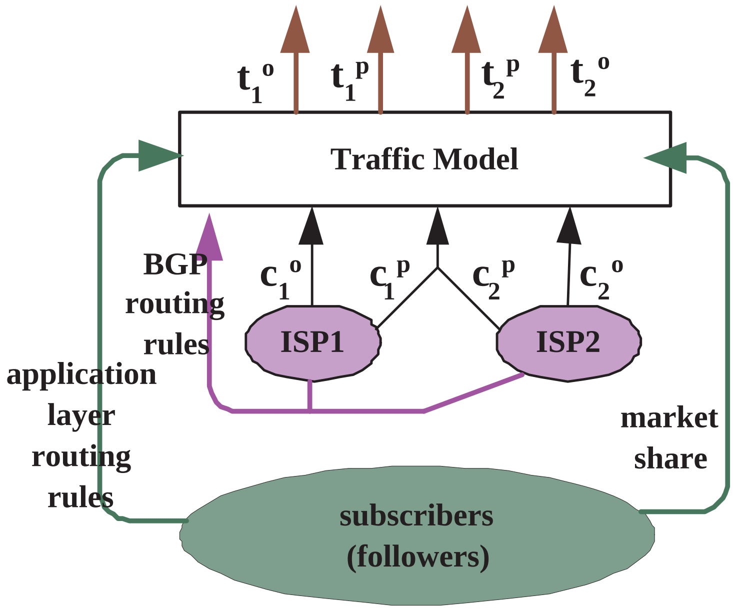

Figure 5.3: Overview of the Traffic Model.

5.4.1 Gravity Modeling Framework

We assume that subscribers in this market are homogeneous, and each subscriber generates traffic demand with an intensity of ρ. Therefore the total demand of ISP_i's subscribers is $D_i = n\rho\alpha_i$. So we have

$$t_i^o + t_i^p + t_i^l = n\rho\alpha_i, \qquad i = 1, 2. \tag{5.12}$$

The task of our traffic model can be seen as to assign the total demand to three different categories - local intra-ISP traffic (t_i^l), peering traffic (t_i^p) and private traffic (t_i^o). We solve this problem using *gravity model* as a reference.

Gravity models are commonly used by social scientists to model the movement of people, goods or information between geographic areas. Generally, those models contain some elements of mass and distance, which lends them to the metaphor of physical gravity as described in Newton's law of gravity. For example, in gravity models for cities, the relative strength of the interaction between two cities might be modeled as directly proportional to the product of the populations and inversely proportional to the distance between the two cities. Recently, researchers start to use gravity model to solve some networking issues such as the estimation of inter-ISP demand matrix [14] and inter-ISP demand matrix [59].

A general formulation of a gravity model is given by the following equation:

$$t_i^j \propto \frac{S_j \times D_i}{F_i^j} \tag{5.13}$$

wherein

1. t_i^j is a variable under study that is related to the flow from j to i.
2. S_j represents the *repulsive* factor that is associated with "leaving" from j;
3. D_i represents the *attractive* factor that is associated with "going" to i;
4. F_i^j is a *distance* factor that is related to the path from j to i.

S_j, D_i and F_i^j should be interpreted appropriately when the gravity model is applied in different contexts under study. In [59], t_i^j is the traffic volume that enters the network at location j and exits at location i. S_j is interpreted as the traffic volume entering the network at location j, and D_i is interpreted as the traffic volume exiting at location i. F_i^j is simply assumed to be a common constant that does not depend on i and j. In [14], t_i^j is the inter-ISP traffic volume from ISP_j to ISP_i. The attractive factor and repulsive factor are interpreted as the population size of one ISP's subscribers ($P_{RA}(\cdot)$) or the size of one

ISP's web contents ($P_{web}(\cdot)$). And F_i^j is the transit quality of the bottleneck ISP in the given path between ISP_i and ISP_j.

The model in [14] is to estimate the intensity of the demand between any pair of ISPs, while the routing of these demand is not considered. We agree that population size and path performance are the most important factors that affect inter-ISP traffic intensity. However, that model cannot serve our purpose to predict the inter-ISP traffic intensity for two reasons. First, ISPs may deploy various approaches to improve the locality of applications and reduce inter-ISP traffic, so subscribers are more likely to connect with other subscribers (or servers) in the same ISP network to retrieve information. Therefore, the attractive factor and the repulsive factor may not be equivalent to the population size of one ISP. Second, we need to predict *traffic* intensity on the inter-ISP links between neighboring ISPs instead of *demand* intensity between any pair of ISPs, thus the routing must be included. Therefore we interpret the gravity model in a slightly different way.

We rewrite Equation 5.13 as follows:

$$t_i^{\#} \ \propto \ S_i^{\#} \times D_i \times G_i^{\#}, \qquad \# = l,\ o,\ p. \tag{5.14}$$

Note it is not a rigorous gravity model since $S_i^{\#}$ depends on the source ISP as well as the destination ISP in order to reflect the locality feature of P2P applications. From Equation 5.12 and Equation 5.14, we have

$$t_i^{\#} = \frac{S_i^{\#} \times G_i^{\#}}{S_i^o\ G_i^o + S_i^p\ G_i^p + S_i^l\ G_i^l} \ \cdot n\rho\alpha_i\ , \tag{5.15}$$

where $G_i^{\#}$ is referred to as *performance factor* (inverse of distance factor). $S_i^{\#}$ and $G_i^{\#}$ will be interpreted in the following subsections.

5.4.2 Interpretation of Performance Factor

The performance factor $G_i^{\#}$ is used to model the effect of application layer routing which is sensitive to link performance.

As we know, the routing of web traffic is controlled by ISPs using BGP. The traffic demand between ISP_i and the Internet is transited via the ISP_i's private link, and the traffic demand between two peering ISPs is transited via the peering link. The routing of web traffic does not usually matter with the inter-ISP link performance.

However, the application layer routing of overlay networks is determined by applications and it is usually sensitive to link performance. Generally speaking, overlay networks attempt to provide "enhanced" services to applications by routing their traffic according to some performance constraints. A degradation in the performance of one path will trigger overlay networks to find an alternate path that satisfies the performance constraints and re-route the traffic accordingly. Although various applications or content distribution networks (CDNs) rely on different peer selection algorithms or service redirection mechanisms, they share the common feature that such mechanisms are more likely to redirect the access to other alternatives if the provisioning is poor across a certain ISP boundary. That is why some ISPs try to use *bandwidth throttling* approach to reduce the inter-ISP traffic intensity.

In our gravity traffic model, we use the performance factor $G_i^{\#}$ to model the sensitivity of the routing to the inter-ISP link performance. We define the *performance sensitivity* function $G(\cdot)$ and let

$$G_i^{\#} = G(c_i^{\#} - t_i^{\#}) \qquad \# = o, p. \tag{5.16}$$

For web traffic we have $G(\cdot) = 1$, which means the routing of web traffic cannot be affected by link provisioning. For overlay traffic, $G(\cdot)$ should be an increasing function, which implies application layer routing always favor the path with good performance. For ease of the presentation, we further assume that $0 < G(\cdot) \leq 1$, and $G(\cdot)$ can be viewed as the *throttling* effect of the link provisioning. We also assume that $G_i^l = 1$ for $i = 1, 2$. This implies that intra-ISP links always have better performance than inter-ISP links.

5.4.3 Interpretation of Repulsive Factor

Subscribers in ISP_i retrieve information from different source ISPs including their own network. Intuitively, larger ISPs are more likely to provide more information for these subscribers since larger ISPs may have more content servers (web traffic) and more peers that are interested in the same files as subscribers in ISP_i (overlay traffic). The repulsive factor $S_i^{\#}$ is used to model the relationship between the amount of one source ISP's contribution and its population size.

In some works peoples assume that the amount of one source ISP's contribution is proportional to its population size. However, many mechanisms deployed by ISPs to improve the locality of traffic make the assumption of proportional contribution unsuitable for current Internet. For example, ISPs deploy various cache servers to make web contents close to their subscribers to reduce transmission cost. After P2P applications become popular, researchers also propose approaches, such as *gateway peers* and *biased neighbor selection* [40], to make more use of local network instead of going to other ISP networks. With the biased neighbor selection approach, the trackers are modified so that the neighbor list of a subscriber A, which is returned by these modified trackers, are more likely to include peers in A's own network, while previous trackers return neighbor list randomly without consideration of peer locality. The inter-ISP traffic intensity can be tuned easily by setting different control parameters with the new tracker algorithm.

An extensive investigation is necessary to understand how these mechanisms quantitatively affect inter-ISP traffic intensities. In this paper, we only provide a framework to model the influence of population size based on a general concept of *caching function*.

We define caching function $r(\alpha)$ as follows. For web traffic, $r(\alpha)$ is the probability that a subscriber's request can be satisfied by its ISP's local network when the subscriber's ISP is with a market share of α. For P2P traffic, $r(\alpha)$ is related to the neighbor selection algorithm of the P2P application, and it can be seen as the average ratio of local peers in the neighbor list of one peer in an ISP network with a market share of α. In other words, $r(\alpha)$ determines the application layer topology of a P2P network, while the transmission rate on links of this network is further determined by $G_i^{\#}$.

$r(\alpha)$ is usually a non-decreasing function of α, the market share (or population size) of the ISP in question, satisfying $0 \leq r(\alpha) \leq 1$. Note, the caching effect can be achieved either by some kind of proxy server (so that other subscribers downloading an object already downloaded previously can be satisfied locally) or by P2P applications with peers in the same ISP serving the role of a proxy server. When $r = 0$, all subscriber requests result in inter-ISP traffic, whereas when $r = 1$, all subscriber requests can be satisfied locally.

Specially, $r(1)$ is the probability that subscribers in this local market connect with servers (peers) in the local market (both ISPs) for content. So $r(1) - r(\alpha_i)$ can be seen as the probability that subscribers in ISP_i connect with servers (peers) in the peering ISP for contents. So the repulsive factor $S_i^{\#}$ is defined as follows:

$$S_i^o = 1 - r(1),$$

$$\begin{aligned} S_i^p &= r(1) - r(\alpha_i), \qquad (5.17) \\ S_i^l &= r(\alpha_i). \end{aligned}$$

If there is no peering link between two local ISPs, the repulsive factor is

$$\begin{aligned} S_i^o &= 1 - r(\alpha_i), \\ S_i^l &= r(\alpha_i). \qquad (5.18) \end{aligned}$$

5.4.4 Inter-ISP Traffic Function

The combination of caching function $r(\cdot)$ and performance sensitivity function $G(\cdot)$ together define a traffic pattern. From Equation 5.15, 5.16 and 5.18, if there is no peering link between two ISPs, we simply have

$$t_i^o = \frac{(1 - r(\alpha_i))\ G(c_i^o - t_i^o)}{(1 - r(\alpha_i))\ G(c_i^o - t_i^o) +\ r(\alpha_i)} \cdot n\rho\alpha_i. \qquad (5.19)$$

For the scenario where there is a peering link between the two local ISPs, we have

$$\begin{aligned} t_i^o &= \frac{r_i^o(\alpha_i)}{r_i^o(\alpha_i) + r_i^p(\alpha_i) + r_i^l(\alpha_i)} \cdot n\rho\alpha_i, \\ t_i^p &= \frac{r_i^p(\alpha_i)}{r_i^o(\alpha_i) + r_i^p(\alpha_i) + r_i^l(\alpha_i)} \cdot n\rho\alpha_i, \qquad (5.20) \end{aligned}$$

where

$$\begin{aligned} r_i^o(\alpha_i) &= (1 - r(1)) \cdot G(c_i^o - t_i^o), \\ r_i^p(\alpha_i) &= (r(1) - r(\alpha_i)) \cdot G(c_i^p - t_i^p), \\ r_i^l(\alpha_i) &= r(\alpha_i). \end{aligned}$$

If overlay routing is load-dependent, Equation 5.19 and 5.20 are fixed point problems and the close-form solution cannot be given easily. In this work, we solve the problems numerically using Matlab.

5.5 Case Studies of ISPs Peering Game

In Section 5.1 we define the ISPs peering game and formulate the game as an optimization problem in Equation 5.6. Since inter-ISP traffic pattern plays an essential role in the peering game, we further present our traffic model to describe different traffic patterns in Section 5.4. Based on the whole framework, now we can analyze the local market and study ISPs' provisioning and peering decisions in the context of different traffic patterns. In this section, we will do a few numerical case studies to gain some insights into how traffic patterns affect ISPs' business strategies.

In all of our case studies, we focus on the same market with the following information. There are $n = 1000$ subscribers in the market, and each subscriber is with a demand of $\rho = 0.001$. The cost function for local ISPs is

$$f(c_i^o, t_i^o) = q_c \times c_i^o + q_t \times t_i^o,$$

where q_t is a constant cost for per unit intensity of traffic and q_c is a constant cost per unit of committed capacity. We further assume $q_t = 80$ and $q_c = 100$ in this paper.

We assume that the subscribers in the local market evaluate ISPs' service performance according to the weighted residual bandwidth of ISP's inter-ISP links as follows:

$$P_i = \frac{t_i^o \times (c_i^o - t_i^o) + t_i^p \times (c_i^p - t_i^p)}{t_i^o + t_i^p}. \tag{5.21}$$

We also assume that the caching function $r(\alpha)$ takes the form of

$$r(\alpha) = \gamma_1 + \gamma_2 \alpha \tag{5.22}$$

in our case studies. γ_1 and γ_2 vary in different traffic patterns, and new caching technologies or new overlay networks can change their values. Let us take early P2P systems as an example. In early P2P systems, trackers randomly return a list of neighbors to each peer. As a result, we can assume all subscribers in one P2P system are homogenous, *e.g.* they make the same amount of contribution to the system. So the caching function for traffic generated by this kind of P2P systems is $r(\alpha_i) = \dfrac{\alpha_i}{1+\alpha_o}$, where α_o is the ratio of peers in the remote Internet to peers in the local market. This is equivalent to our model with $\gamma_1 = 0$ and $\gamma_2 = \dfrac{1}{1+\alpha_o}$.

For any traffic pattern, the caching effect should always be non-negative and should increase with the population size. It implies that $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$. In addition, the marginal cost of a new coming subscriber should be positive in a local ISP, which means that the inter-ISP traffic intensity should increase with the ISP's population size. So we have $\dfrac{\partial(1-r(\alpha))\alpha}{\partial\alpha} \geq 0$, which is equivalent to $\gamma_1 + 2\gamma_2 \leq 1$.

ISPs can determine $r(\alpha)$ for the mixed traffic in their market through various measurement technologies, and then apply our game theoretic framework to study provisioning and peering strategies in their market. Researchers can propose reasonable models for any inter-ISP traffic pattern generated by individual popular system. Our game theoretic framework can be easily applied to study the economic implications of the popular system on ISPs' business.

We fix ISP_2's private capacity as $c_2^o = 0.6$ and conduct numerical studies to help ISP_1 understand the market and make proper decisions.

5.5.1 Provisioning Strategy: Web Traffic

In this case study, we focus on web traffic, whose routing is controlled by ISPs using BGP. We assume the dominant traffic in current market is web traffic with $\gamma_1 = 0$ and $\gamma_2 = 0.1$. In addition, there will be a new technology that can improve the locality (caching effect) of those web traffic and then $\gamma_1 = 0.05$ and $\gamma_2 = 0.2$.

ISP_1 needs to decide two things, *e.g.* how much capacity it needs to buy and whether to peer with ISP_2 or not. The Nash Equilibria under the two traffic patterns in the context of different provisioning and peering strategies are plotted in Figure 5.4 (α_1^*) and Figure 5.5 (p_i^* and $\mathcal{R}_i^*$). Note that each point in the figures shows the Nash Equilibrium of one game, which is defined by a particular provisioning strategy (c_1^o) and a particular peering strategy under the traffic pattern with a particular caching function.

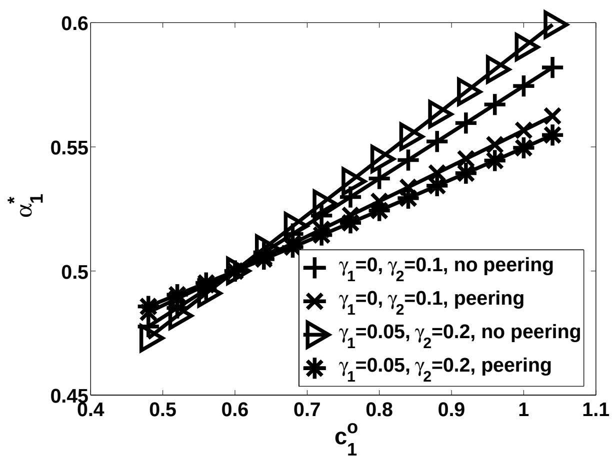

Figure 5.4: α_1^* in Different Scenarios (Web Traffic).

In this market, ISP_1 can induce more subscribers by increasing the capacity of its private links. After c_1^o increases, ISP_1 can charge its subscribers more, while ISP_2 has to lower its service price in order to maximize its profit. In this particular market, ISP_1's profit increases after ISP_1 buys more private capacity, while ISP_2's profit decreases because its price decreases.

Without peering, the market share of the larger ISP (with more private capacity) would increase after the new technology

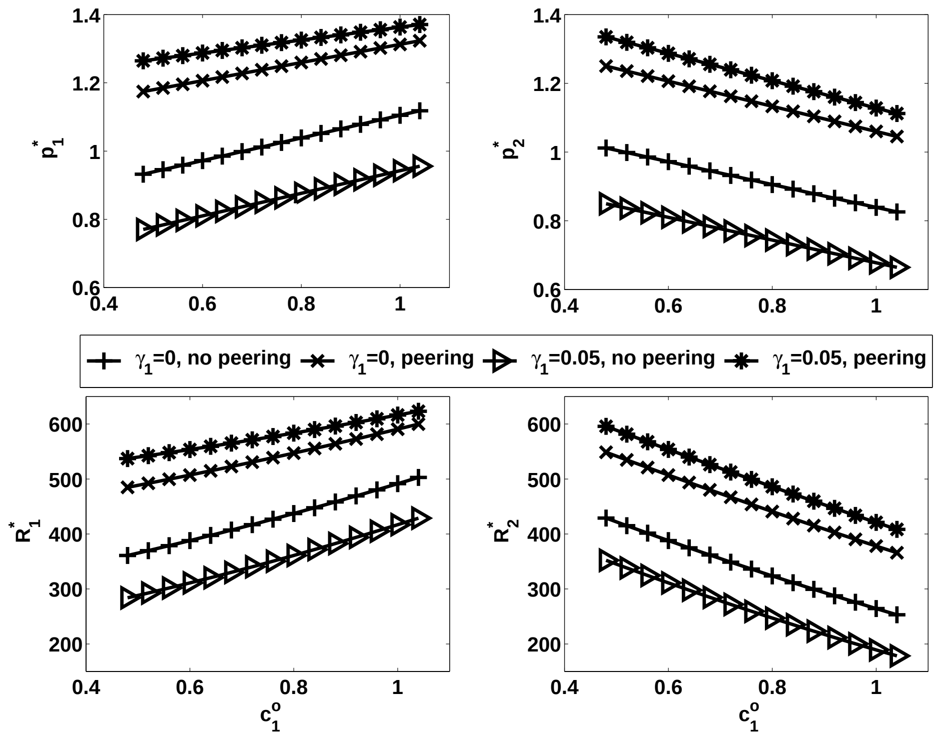

Figure 5.5: p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Web Traffic).

is deployed (such as the deployment of CDNs). It implies that the larger ISP can benefit more from the new technology, which is consistent with *economy of scale*. The new technology improves the caching effect, so the marginal cost of one subscriber is lowered, which means the competition for subscribers between two ISPs would be fiercer than before. Both ISPs lower its service price to induce subscribers, therefore subscribers also benefit from the technology. However, both ISPs' profits decrease due to the fiercer competition.

Peering is always bad for the ISP with more private capacity when only market share is concerned. This results in that large ISPs do not want to peer with small ISPs. However, both ISPs can charge their subscribers more, thus make more profit after peering. This might explain why ISPs with comparable size would like to peer with each other.

Market share is less sensitive to the provisioning of two ISPs' private links after peering. Peering even reverses the influence of the new technology. After peering, the appearance of the new technology would increase the market share of the small ISP. Peering combines two ISP networks into one single network, thus two ISPs enjoy the same economy of scale. In addition, subscribers even need to pay more after the deployment of the new technology with peering, while without peering the Internet access cost would be lowered after the deployment of the new technology.

5.5.2 Provisioning Strategy: Overlay Traffic

In the second case study we focus on P2P overlay traffic. We assume that the performance sensitivity function $G(\cdot)$ for P2P overlay traffic takes the following form:

$$G(\delta) = \begin{cases} \frac{1}{2}(\frac{\delta}{c_m})^{2h_m c_m} & \text{if } \delta < c_m, \\ \\ 1 - \frac{1}{2}(\frac{c_m}{\delta})^{2h_m c_m} & \text{if } \delta \geq c_m, \end{cases} \tag{5.23}$$

where $h_m c_m > 1/2$.

$G(\delta)$ is to measure the response of overlay routing to the link performance evaluated by residual capacity of the link. We define the performance sensitivity function in this way because the above function satisfies the following features.

The performance factor approaches to zero as the residual capacity of the link approaches to zero, *i.e.* $\lim_{\delta \to 0} G(\delta) = 0$. The traffic on a link with large residual capacity would not be throttled by the link performance, *i.e.* $\lim_{\delta \to \infty} G(\delta) = 1$. The throttling effect is diminishing as the residual capacity increases, *i.e.* $G'(\delta) > 0$. More importantly, the changes of traffic intensity (equivalent to G') would be slower and slower as the residual capacity approaches to zero or infinity, *i.e.* $\lim_{\delta\to 0} G''(\delta) = 0$ and $\lim_{\delta\to\infty} G''(\delta) = 0$. In other words, further increasing the inter-ISP link capacity will not induce a lot of traffic after the residual capacity δ exceeds a threshold, and further decreasing the inter-ISP link capacity will not reduce the traffic intensity a lot if δ has been below a certain level.

h_m and c_m are parameters which can be tuned to model traffic patterns generated by different overlays. We believe that $c_m = G^{-1}(1/2)$ is also related to the average capacity of intra-ISP links. h_m reflects the sensitivity of the overlay routing to link performance when δ is around c_m, and greater h_m implies more sensitive overlay routing. Figure 5.6 is an illustration of $G(\delta)$ with different parameters.

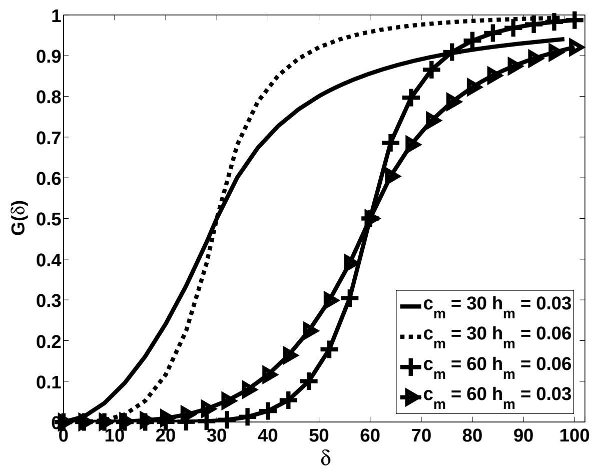

Figure 5.6: Performance Sensitivity Function $G(\delta)$.

In this case study, the first traffic pattern under study is with $\gamma_1 = 0$, $\gamma_2 = 0.1$, $c_m = 1.4$ and $h_m = 0.4$. People may propose various ways to improve locality and routing sensitivity of overlay traffic. Our second traffic pattern is used to model this trend by setting $\gamma_1 = 0.1$, $\gamma_2 = 0.2$, $c_m = 1.4$ and $h_m =$

0.6. Similar as the first case study, we solve all the games with different provisioning profiles, peering strategies and different traffic patterns. The result Nash Equilibria of these games are plotted in Figure 5.7 (α_1^*) and Figure 5.8 (p_i^* and $\mathcal{R}_i^*$).

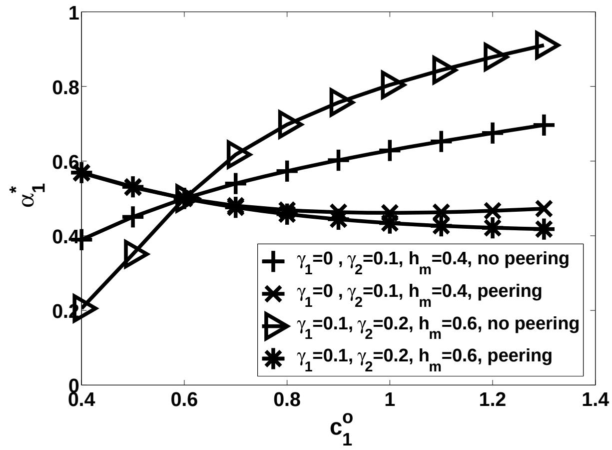

Figure 5.7: α_1^* in Different Scenarios (Overlay Traffic).

When two ISPs do not peer with each other, increasing private capacity results in more subscriber share. After peering, however, increasing the provisioning reduces its own market share and increases its competitor's profit. It might be caused by overlay routing, which renders more difficulties for ISPs to achieve their particular goals. ISP_1 increases c_1^o, and it makes that more peers in ISP_1 retrieve information from outside peers instead of from peers in the local market due to performance-sensitive routing. Therefore, ISP_1's inter-ISP intensity increases, which may partly counteract the performance improvement from the increased provisioning. At the same time, ISP_1 needs to pay more because of the increased capacity and increased inter-ISP traffic intensity. As a result, ISP_1 has to increase the price although its performance improvement is limited, which may drive away its subscribers. Comparing with the market with only web

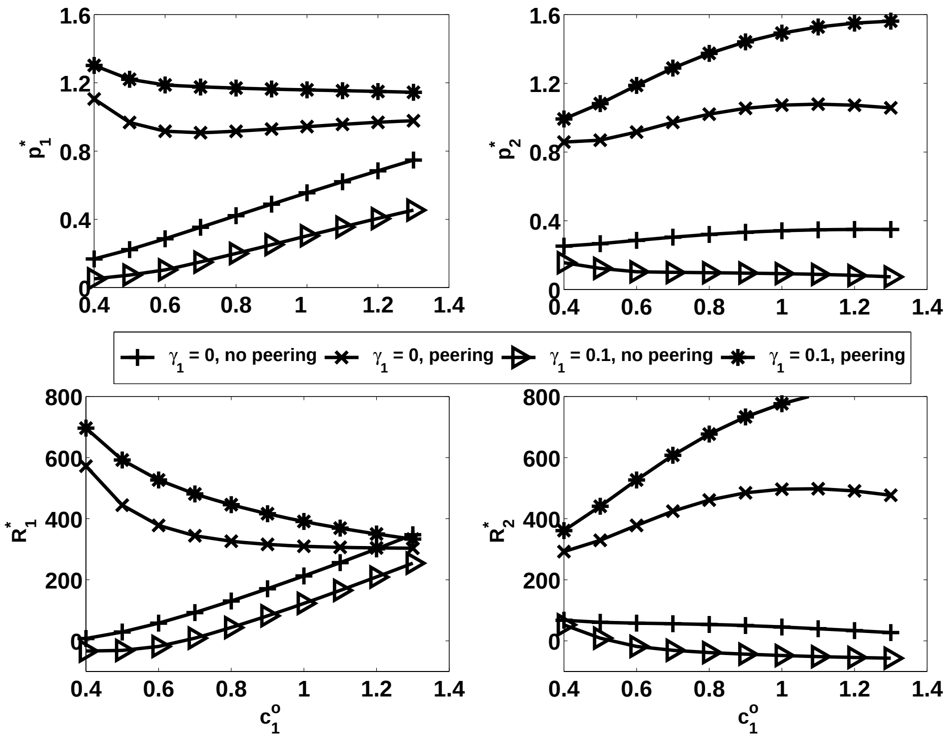

Figure 5.8: p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Overlay Traffic).

traffic, it is more difficult for two ISPs to make a free peering agreement due to application layer routing.

Figure 5.7 and 5.8 show that it might be a good strategy for one ISP to reduce its private link provisioning under the peering scenario. The ISP's market share does not decrease and its profit increases. This may explain why some ISPs begin to use *bandwidth throttling* to contain inter-ISP P2P traffic intensity. Obviously, with peering relationship, the traffic pattern under study discourages network growth and service improvement under current Internet architecture and business model.

Before peering, new technologies that improve the locality and the performance sensitivity of overlay applications would also make the competition between two ISPs fiercer. ISPs reduce their price to induce more subscribers because subscribers themselves are serving as information providers and thus become more valuable for ISPs. After peering, both ISPs can make use of all peers in the whole market, so it may not be necessary for them to reduce the service price to attract subscribers.

In terms of market share, peering is always bad for the ISP with larger capacity. The negative influence is more apparent in a market with overlay traffic than a market with web traffic. But after peering two ISPs can increase their prices at the same time, thus make more profit. To some extent, we can say that two ISPs are partly cooperating with each other under the peering scenario.

5.5.3 Provisioning Strategy: Traffic Transition

Several years ago, web traffic was the dominant traffic in the Internet, while in recent years, overlay traffic has been more and more popular. In this study, we focus on this transition the current Internet is going through. Particularly, we study four scenarios: from the extreme case of 100% web traffic to the middle case of 60% web traffic (and 40% overlay traffic), and then the middle case of 30% web traffic to the extreme case of 100% overlay traffic.

In this case study, we still assume that $c_2^o = 0.6$, $\gamma_1 = 0.1$, $\gamma_2 = 0.2$, $c_m = 1.4$ and $h_m = 0.4$. We also assume there is a peering link between two local ISPs. We solve all the games with different provisioning profiles and different traffic patterns, and the result Nash Equilibria of these games are plotted in Figure 5.9 (α_1^*) and Figure 5.10 (p_i^* and $\mathcal{R}_i^*$).

We see that increasing private capacity results in smaller advantages (even disadvantages) as the ratio of overlay traffic increases. When there is only web traffic, increasing private capacity can induce more subscribers. However, when the ratio of overlay traffic exceeds a certain threshold (around 60% in this

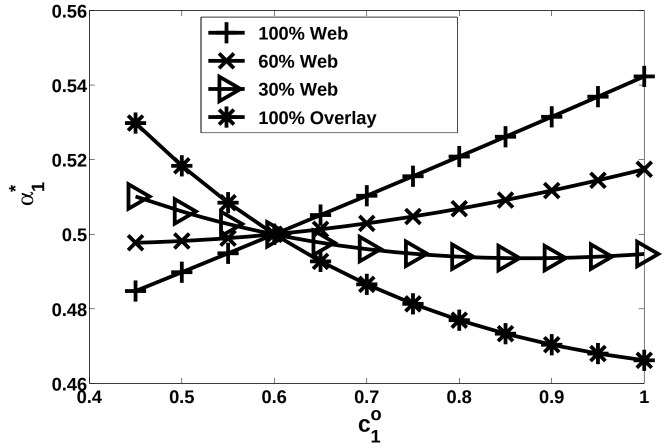

Figure 5.9: α_1^* in Different Scenarios (Peering).

case study), increasing private capacity reduces its own market share and sometimes even increases its competitor's profit, which means the side effect due to the free ride phenomenon has overcome the transmission cost saving from local peering. Figure 5.9 also shows that peering brings more harm to the large ISP when there is more overlay traffic. For example, assume the capacity of ISP_1 is 0.8, which is larger than ISP_2's private capacity, we can find that ISP_1's market share keeps decreasing as the ratio of overlay traffic increases.

5.5.4 Peering Strategy

Now let us look into the peering strategy. Peering strategy is more than deciding peering or not. ISPs can control the peering capacity between them to achieve their individual goals. In some networks, peering capacity has significant influence on ISPs' business.

In this case study, the local market is the same as before. We further assume ISP_1's private capacity is also fixed and $c_1^o = 1$. We will look at the influence of peering capacity in the context

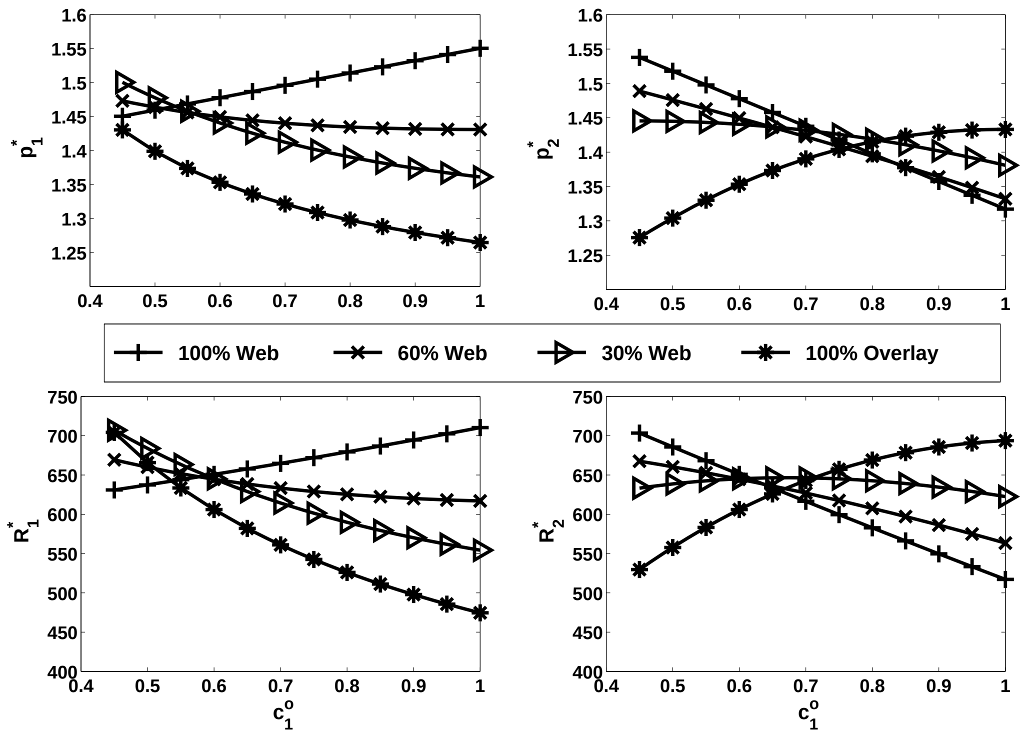

Figure 5.10: p_i^* and $\mathcal{R}_i^*$ in Different Scenarios (Peering).

of three traffic patterns. These three traffic patterns are with the same caching function where $\gamma_1 = 0$ and $\gamma_2 = 0.1$. The first traffic pattern is web traffic. The other two traffic patterns are overlay traffic patterns with $c_m = 1.4$ and $h_m = 0.4$ or 0.6. The Nash Equilibria $(\alpha_1^*, p_i^*, \mathcal{R}_i^*)$ of peering games under different peering capacities are plotted in Figure 5.11 and Figure 5.12.

For web traffic, the routing is determined by ISPs, and only local traffic can go through the peering link. So the traffic intensity on the peering link is always below a certain value. α_1^* would not be reduced a lot by increasing peering capacity. However, overlay traffic is not under the control of the peering agreements. Overlay applications can continue moving traffic to the path with good performance, so increasing peering capacity exasperates the free riding of the ISP with worse performance. It

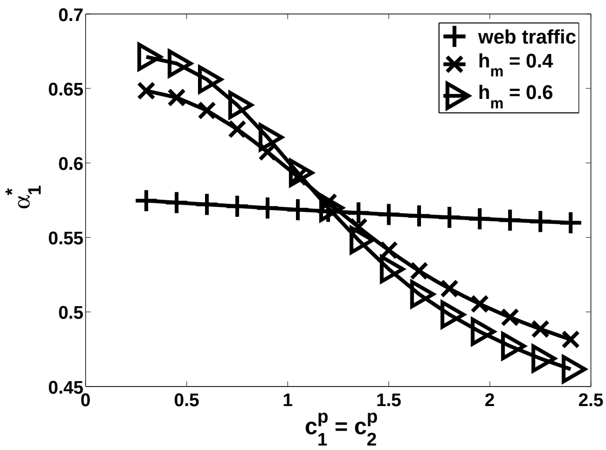

Figure 5.11: α_1^* under Different Peering Capacities.

may reduce the market share of the larger ISP (α_1^*) a lot when overlay traffic dominates. In addition, the negative influence would be larger if the overlay traffic is more performance sensitive. Therefore, with overlay traffic, ISPs should be very careful when making the peering decisions.

We also notice that ISPs' prices keep increasing as two ISPs increase their peering capacity. It reveals that two ISPs form a tighter coalition and they have more bargaining power to subscribers after their peering capacity increases, although the peering benefit distribution is not fair, which can be seen from the change of the market share distribution shown in Figure 5.11.

5.5.5 Discussion on Subscribers' Choices

In the above analysis, two ISPs are able to provide satisfactory service for all subscribers in this market (or Internet access service is necessary for subscribers in this market), so what the subscribers need to decide is which ISP to subscribe. This market can be thought as saturated, where all potential subscribers have joined the Internet. These subscribers always choose the

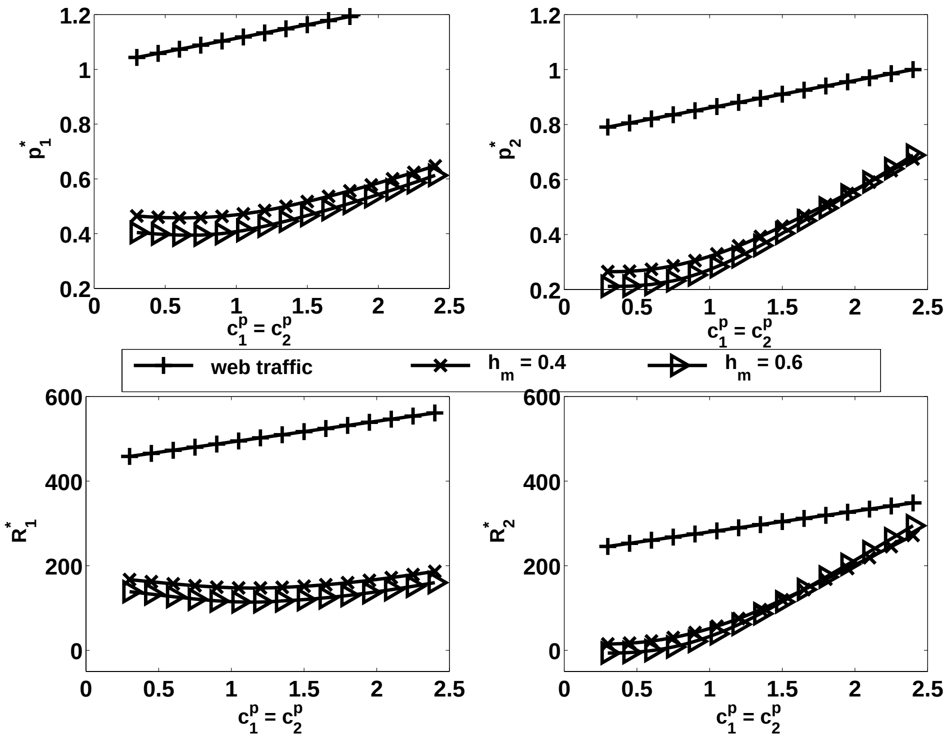

Figure 5.12: p_i^* and $\mathcal{R}_i^*$ under Different Peering Capacities.

ISP with high service utility, thus the utility of subscribers in two ISPs evaluate to the same value when the market becomes stable. In this model, two ISPs are competing with each other, and both of them have incentives to lower the price to attract subscribers from the other ISP. As a result, subscribers enjoy reasonable prices.

We can allow that some of n potential subscribers do not subscribe to any ISP to model an expanding market. In this market two ISPs are not able to provide satisfactory service for all potential subscribers. One subscriber always chooses the ISP with high service utility, and the service utility must be non-negative. Otherwise, the subscriber does not subscribe the Internet access service. With this assumption, we still have the following equation for the Nash Equilibrium:

$$\mathcal{U}_1(p_1^*, \alpha_1^*) = \mathcal{U}_2(p_2^*, \alpha_2^*).$$

There is one more condition for the Nash Equilibrium as follows

$$\mathcal{U}_1(p_1^*, \alpha_1^*) \geq 0, \qquad \mathcal{U}_2(p_2^*, \alpha_2^*) \geq 0. \tag{5.24}$$

Let us assume that some users do not choose any ISP when the market becomes stable, *i.e.* $\alpha_1^* + \alpha_2^* < 1$ at the Nash Equilibrium point. Obviously we have

$$\mathcal{U}_1(p_1^*, \alpha_1^*) = 0, \qquad \mathcal{U}_2(p_2^*, \alpha_2^*) = 0. \tag{5.25}$$

Otherwise subscribers would choose the ISP with positive utility. In this case, two ISPs are not competing with each other directly since there are "free" potential subscribers which do not belong to any ISP.

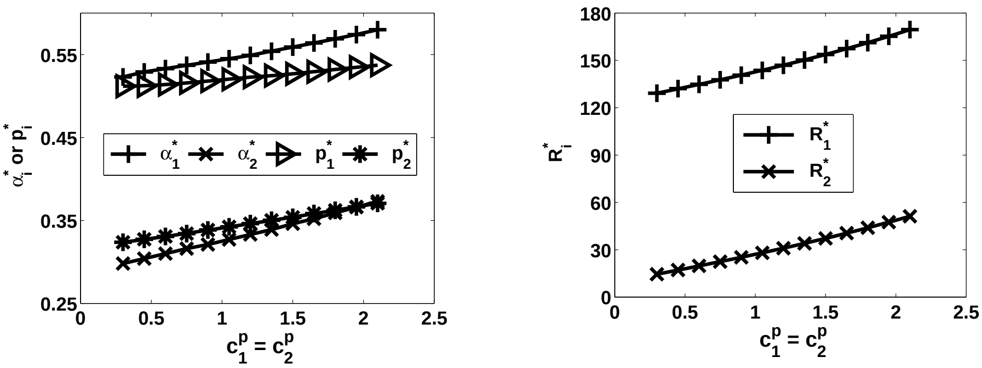

Figure 5.13: Some Subscribers Do Not Choose Internet Access Service.

We conduct the numerical study under the new assumption for the web traffic in the third case study [1]. The result is shown in Figure 5.13.

We have two observations from Figure 5.13. First, the two ISPs' prices are lower than the result in Figure 5.12. In order

[1] We ignore the two overlay traffic patterns since two ISPs provide positive utility for all subscribers in the market at Nash Equilibria points under those two overlay traffic patterns.

to avoid some subscribers leaving the Internet, both ISPs lower their prices to attract potential subscribers. However, the same as in a saturated market, the two ISPs' prices keep increasing as they increase the peering capacity. Second, the market shares of both ISPs are increasing with the peering capacity, but the the market share of the smaller ISP increases more quickly than the larger ISP, which shows that the smaller ISP benefits more from the peering relationship than the larger ISP.

This case study is for an expanding market, where two ISPs cannot provide satisfactory service for all potential subscribers; while the first three case studies can be viewed as for a saturated market, where all subscribers have joined the Internet and two ISPs are competing with each other directly.

5.6 Related Work

Peering strategy is important for ISPs' business. Based on interviews with several hundred ISPs Peering Coordinators, William Norton presents several important Internet operation research papers [54] [55] [56] [57] [58]. These papers describe various issues in ISPs' peering from practical point of view, and they are very important for academic researchers to understand peering in real world.

In [54], the author introduces two models for ISP interconnection: the exchange-based interconnection model and the direct circuit interconnection model. He compares these two options from a technical and business (financial) standpoint by the introduction and application of the Neutral Internet Business Exchange as the basis for the exchange-based model comparison. Real world costs and revenue projections are used in his financial model to quantify the cost savings and revenue generated by participants in both models. As far as we know, it is the first document to conduct economic analysis for ISPs' peering.

[55] is an interesting paper where the author details the ISP peering decision-making process from selection of potential peers through implementation. The author highlights three distinct phases in the peering process: Identification (Traffic Engineering Data Collection and Analysis), Contact & Qualification (Initial Peering Negotiation), and Implementation Discussion (Peering Methodology). The first phase identifies the who and the why, while the last phase focuses on the how. This paper also includes the description of a Peering Simulation Game that can play out peering negotiations.

In [56], the author builds upon the foundation of peering knowledge and present tactics that Peering Coordinators have used to obtain peering where they otherwise might not have been able to obtain peering. They have identified 19 specific maneuvers that vary from mundane to the clever, from merely deceptive to manipulative. In sum, these tactics represent the "Peering Playbook", the current "Art" of Peering.

[57] introduces the tools and analysis typically used by Peering Coordinators. The most important contributions in this paper are "Peering Break Even Analysis Graph" and the financial model. Using the analysis graph, we can find the "Peering Breakeven Point" where ISPs are financially indifferent between peering and simply sending all traffic through its upstream ISP. Once traffic volume between the ISP and the peering population reaches the breakeven point, ISPs start saving money peering. The "Peering Risk" is the range of traffic exchange where an ISP fails to exchange enough traffic with the other ISPs at the exchange point to offset the cost of peering. This paper is a business case for Internet Service Provider peering based on real world practices and market prices.

About seven years ago, a broad set of ISPs that were once focused only on growing their market share (at any cost) now are bending down to shave pennies off of their cost structure. [58]

focuses on this Evolution of the U.S. Peering Ecosystem. Several key players have entered into the Peering Ecosystem and caused a significant disruption to the Ecosystem. Peer-to-Peer application traffic has grown to represent a significant portion of their expense. The author describes five major events (such as collapse of the Telecom Sector and emergence of P2P applications) and three emerging evolutions (such as cable companies are peering and content companies are peering) in the Peering Ecosystem. They also present a simple mathematical Internet Peering Model that can be used to demonstrate this Peering Ecosystem evolution. While not complete or by any means precise, it does demonstrate the affect of these disruptions in the Peering Ecosystem.

Over the last few years, overlay networks have become increasingly popular as a promising platform to provide customizable and reliable services at the application layer. As an important type of overlay networks, the rise of P2P applications in broadband networks also draws a lot of attentions. It has been already reported and discussed academically as well as in popular press [16] [24]. The emergence of various overlay networks raises many technical challenges and business issues for ISPs.

Most of current works related to overlay networks focus on technical challenges. For example, there are various efforts in trying to characterize P2P traffic and studying its caching behavior [23] [47]. Researchers are also making efforts to improve the performance of overlay networks [4] [40], and the problematic interactions between IP networks and overlay networks are examined in [45]. In [34], the author point out that overlay networks attempt to take control over routing in the hope that they might achieve better performance and illustrate how an uncoordinated effort of the two layers to recover from failures may cause performance degradation for both overlay and non-overlay traffic. It also reveals that current traffic engineering techniques are inadequate to deal with emerging overlay network services.

Due to the inter-dependency nature of the economic principles and the technical architecture of the Internet operations, there has been many works applying economic and game-theoretic analysis to study Internet protocol and operational issues. In [44], the authors give an interesting analysis of how transit and customer prices are set in a network with multiple ISPs competing for the same customers. They also examine the existence of equilibrium price strategies and show how positive profit can be achieved using threat strategies. In [36], the problem of where a pair of ISPs will make their peering connections so as to minimize their own transit traffic is formulated and studied by a game-theoretic model. [9] introduces a novel non-cooperative game model to characterize the caching problem among selfish servers without any central coordination and analyze replication of resources by these server nodes. In addition, authors in [53] explore the interactions between the routing behaviors of overlay networks and the underlay network, revealing the properties and implications of inefficient routing equilibria. All of these studies apply economic and game-theoretic analysis to networking problems and generate nice insights to the design of networking protocols and operational principles.

Some researchers work on the analysis of peering strategy based on their economic models and traffic models. In [28] [57], the authors predict when these ISPs would use no-payment peering rather than customer-provider peering for their local traffic. In [57], the traffic volume between two ISPs should be known in advance. In [28], the author proposed a simple traffic model to derive the traffic volume based on the known market shares of two ISPs. However, only the effect of traditional web and email traffic are considered in the traffic model. Some new network traffic, such as traffic generated by P2P applications, may give a totally different impact on peering decisions.

The study on economic implications of overlay traffic, especially P2P traffic, has so far received less attention. As far as we know, we are the first to study the impact of application layer routing in overlay networks on ISPs' peering and provisioning strategy.

5.7 Summary

Overlays gain widespread significance as a method for end users and third parties to affect routing decisions and they become vehicles for contention over the routing decision between ISPs and end users. The use of overlays may create substantial effects for ISPs, and potentially for the overall Internet architecture. What are the implications of such overlay network traffic on ISPs' peering? Can ISPs still rely on Internet's policy-based routing protocol (BGP) to manage their traffic, or will they need new strategies or new Internet business model in view of the overlay traffic?

In this chapter, we study a local market with two ISPs competing with each other for market share of subscribers, assuming rational ISPs and subscribers. We formulate the interaction among ISPs and subscribers in this market as a two-stage game, where subscribers picking ISPs, and ISPs setting prices, making provisioning and peering decisions. Based on this game formulation, we study the influences of different traffic patterns on the Nash Equilibrium of the market. The goal of this analysis is to establish an abstract and quantitative framework that brings out the important factors that affect ISPs' peering.

We provide a general inter-ISP traffic model to define traffic patterns based on caching function and performance sensitivity function. Our numerical study focuses on two traffic pattern families. One is for the traditional web traffic, whose routing is determined by ISPs. The other is for overlay traffic, which represents the traffic whose routing is reactive to the network performance.

Then we conduct numerical studies to understand the influences of these two traffic patterns on the peering game in the market. The result reveals some important observations: (1) while *economy of scale* is the predominant property of the competitive ISP market, overlay traffic may introduce unfair distribution of peering benefit (*i.e. free-riding*); (2) the large ISP can restore more fairness by reducing its private capacity (*bandwidth throttling*), which has the drawback of hurting business growth; (3) ISPs can reduce the level of peering (*e.g.* by reducing peering bandwidth) to restore more fairness, but this has the side-effect of also reducing the ISPs' collective bargaining power towards subscribers.

The insights from this work is useful for further study of ISP peering practices, and considerations for different interdomain routing mechanisms and other tools to support these new peering practices. From a single ISP's perspective, our methodology can also be used to build a decision support system to evaluate various peering and provisioning strategies.

□ **End of chapter.**

Chapter 6

Conclusion

The Internet is a "network of networks" operated by many ISPs who interconnect with each other at IXPs under different business relationships (*e.g.* peering relationship or transit relationship). These ISPs compete with each other on price, performance, reliability, *etc.*, but at the same time they must cooperate with each other to provide global connectivity to all other attachments on the Internet.

In running its business, each ISP should select its provider ISPs and peering ISPs based on its economic considerations. The ISP needs to further negotiate with its neighboring ISPs on where to interconnect and the capacity of all inter-ISP links. After that, the ISP should try to make best use of these inter-ISP links to satisfy its customers and make profits through various traffic engineering techniques. In this thesis, we focus on ISPs' peering strategy and interdomain traffic engineering. These issues are two important aspects of ISPs' business and have significant implications on the Internet architecture.

Our study on interdomain traffic engineering focuses on AS Path Prepending (ASPP), a popular way for ISPs to implement inbound traffic engineering. In order to improve the current situation that ISPs often practise this approach in a trial-and-error basis, we propose a greedy algorithm to help ISPs perform it systematically and efficiently. Our simulations show that this algorithm converges quickly to satisfactory results.

ISPs in the Internet must cooperate with each other to transit packets from sources to destinations. At the same time, each ISP has its own objective and interest, so that the local traffic engineering practises of ISPs may be based on different preferences on the routing of packets. We illustrate some fundamental issues of this decentralized traffic engineering in the Internet based on an abstract model where ISPs perform traffic engineering for their individual load balance. First, the traffic engineering practise to improve local performance index may make the global performance index worse. Second, the uncoordinated local behaviors by multiple ISPs can result in routing instability.

To see if such prepending instability really occurs in the Internet, we analyze the routing tables from Route Views to study the dynamic prepending behavior of ISPs. Based on a random picked snapshot, we extract 229 prefixes whose routes satisfy the CRR condition for further analysis. For each extracted prefix, we analyze its routes over a period of several months to see how different ISPs change their ASPP actions. We also look into the corresponding routing tables to find the best route and see how the prepending actions affect the best route selection. Our preliminary study reveals a real-world pathologic case of prepending instability. In this pathologic case, AS_a's prepending length changes from 0 to 3 and back many times, and AS_b's prepending length changes from 2 to 3 back and forth. As a result of these repetitive prepending actions, the best route for this prefix involves an oscillation.

Routing instability can result in poor end-to-end network performance and degrade the overall efficiency of the Internet infrastructure. Therefore, we present some simple guidelines for network operators to avoid the routing instability caused by decentralized traffic engineering.

The above discussion on traffic engineering is based on a "fact" that ISPs can completely control the routing of all packets by tuning their routing policies. Also because of this "fact", ISPs' settlement was what they expected from their negotiated peering agreements, which made ISPs' peering strategy relatively simple. However, over the past several years, we have seen the emergence of numerous types of "overlay" networks in the Internet to enhance or modify the basic functioning of traffic handling within the Internet. Because the routing of overlay traffic at application layer is controlled by overlay networks, ISPs lose the routing control of some interdomain traffic flows and that "fact" is no longer true. With application layer routing, one ISP may provide unintended transit service for its interconnected ISPs and it is difficult to predict the settlement between ISPs only based on their business agreements. This new phenomenon upsets the traditional business model and makes ISPs' peering strategy more complicated.

In the second part of this thesis, we focus on the economic implications of different inter-ISP traffic patterns on ISPs' peering strategy. In Chapter 4, we conduct an economic analysis for an overlay streaming network and present some insights on the issues brought by overlay networks. The routing of the overlay is modelled as an optimization problem to capture the important difference between how overlay traffic and tradition web traffic are routed in the Internet - the routing of traditional web traffic is determined by ISPs according to ISPs' interests, while the routing of overlay traffic is controlled by applications for applications' interests. Therefore in our routing model, the routing of overlay streaming traffic is to optimize the performance of transferring each byte of overlay demand.

In this analysis, we deploy the break-even price as the metric to evaluate the economic efficiency of one ISP. We conclude that the ISPs must control both their provider link capacity as well as free-peering capacity they provide to other peers to maintain certain parity between different incoming link capacities among all ISPs. Only when such parity is maintained, ISPs can sustain their relative economic positions despite of application layer routing.

There are two limitations with the analysis in Chapter 4. First, it cannot be applied directly on other overlay traffic types since we only focus on one particular overlay traffic model. More importantly, that analysis is conducted in a static market, where both the market share distribution and the provisioning of all links are fixed. As a result, that economic analysis does not take the response of subscribers into consideration, and the evaluation based on break-even price only reveals the direction of the influence of overlay traffic at a particular point of the market, but cannot capture the stable state (equilibrium) of a dynamic market.

We improve this static economic analysis in Chapter 5. The dynamic market of two ISPs is formulated as a multi-leader-follower game to capture the Nash Equilibrium of the routing tussle. This game-theoretic framework captures the decision of the major players of the Internet marketplace - the subscribers (users) and the ISPs playing different roles in providing transit service, and provides Nash Equilibrium as an output to represent the steady state of a dynamic market.

We adopt a gravity traffic model as a simple and intuitive way to model the inter-ISP traffic patterns generated by different applications. In this model, traffic flows according to the number of subscribers in each ISP engaged in the same application at the same time, modulated by performance influence which leads to specific route selection by overlay networks. In other words, an ISP with more subscribers tends to attract (or contribute) more traffic, with the throttling of the performance of links connecting this ISP to its neighbors. By adjusting its parameters, such a gravity model provides a rational approximation of different inter-ISP traffic patterns.

Based on this framework, including the game theoretic formulation and the traffic model, we provide some insights on peering strategy under overlay traffic scenarios: (1) while *economy of scale* is the predominant property of the competitive ISP market, overlay traffic may introduce unfair distribution of peering benefit (*i.e. free-riding*); (2) the large ISP can restore more fairness by reducing its private capacity (*bandwidth throttling*), which has the drawback of hurting business growth; (3) ISPs can reduce the level of peering (*e.g.* reducing peering bandwidth) to restore more fairness, but this has the side-effect of also reducing the ISPs' collective bargaining power towards subscribers. From a single ISP's perspective, our methodology can also be used to build a decision support system to evaluate various peering and provisioning strategies under different interdomain traffic patterns.

The network architecture, along with the traffic engineering policies that ISPs use, are based on certain assumptions about how their customers and traffic behave. Overlay networks could call into question some of these assumptions, thus rendering it more difficulties for ISPs to achieve their goals. The implications of overlay traffic on peering strategy have not been investigated before. We believe that this study is helpful for ISPs to have better knowledge about various traffic patterns to cope with some technical issues and business issues, and the insight of this work may lead to further examination of the Internet architecture and interdomain routing mechanisms. Therefore, we believe this research can potentially have high practical as well as theoretical significance.

□ **End of chapter.**

Appendix A

Solving the Optimization Problem

In this chapter, we solve the optimization routing problem defined in Equation 4.6 and derive the close form routing solution. First, we can rewrite Equation 4.6 as follows:

$$\begin{array}{rll}
\mathbf{min}_{y_1,y_2} & f(y_1,y_2)=\alpha_2(y_1-y_1^{'})^2+\alpha_1(y_2-y_2^{'})^2, & \\
\textbf{subject to} & g_1=-y_1\le 0 & g_2=-y_2\le 0 \\
& g_3=1-c_2^o-y_1\le 0 & g_4=1-c_1^o-y_2\le 0 \\
& g_5=y_1-1\le 0 & g_6=y_2-1\le 0 \\
& g_7=y_1-c_2^p\le 0 & g_8=y_2-c_1^p\le 0 \\
& g_9=y_1+y_2-1\le 0 & \\
\textbf{where} & y_1^{'}=\dfrac{c_2^p-c_2^o+2}{4}, & y_2^{'}=\dfrac{c_1^p-c_1^o+2}{4}.
\end{array}$$

The lagrangian function is $L=f+\sum_{i=1}^{9}\mu_i g_i$ and we can write the KKT conditions as follows:

$$\frac{\partial L}{\partial y_i}=0 \tag{A.1}$$

$$\mu_i g_i=0 \qquad i=1\ldots 9 \tag{A.2}$$

$$u_i\ge 0 \qquad i=1\ldots 9 \tag{A.3}$$

$$g_i\le 0 \qquad i=1\ldots 9 \tag{A.4}$$

Let us further analyze the KKT condition. Equation A.2 includes $2^9 = 512$ cases ($\mu_i = 0$ or $g_i = 0$). For example, consider the case that $u_i = 0$ for all $i = 1 \ldots 9$, it means we are guessing all constraints are not active. Upon using other KKT conditions, we get:

$$y_1 = y_1^{'}, \quad y_2 = y_2^{'}, \quad g_i \leq 0.$$

Obviously, if $(y_1^{'}, y_2^{'})$ satisfies the condition $g_i \leq 0$ (*i.e.* the center of the ellipse is in the feasible region), the minimum point is this center $(y_1^{'}, y_2^{'})$. Otherwise, we have violated the KKT conditions and hence our guessing is incorrect. Although this case is easy to understand, applying the lagrangian method based on this approach of guessing directly is not a solution technique since there are too many cases.

In the following discussion, we classify these 512 cases in three categories, as stated in Subsection 4.3.2, to simplify the analysis.

A.1 No Peering: $y_1^* = 0$ and $y_2^* = 0$

Let us look at the conditions for the point $(0, 0)$ to be a KKT point. Since y_1 and y_2 are symmetric in our optimization problem, we only analyze the condition related with y_1 in the following discussion.

$$\begin{aligned}
u_i g_i = 0 \;&\Rightarrow\; u_5, u_9 = 0, (c_2^o = 1 \text{ or } u_3 = 0), (c_2^p = 0 \text{ or } u_7 = 0); \\
g_i \leq 0 \;&\Rightarrow\; c_2^o \geq 1; \\
\frac{\partial L}{\partial y_1} = 0 \;&\Rightarrow\; 2\alpha_2(-y_1^{'}) - u_1 - u_3 + u_5 + u_7 + u_9 = 0 \\
&\Rightarrow\; u_1 = \frac{\alpha_2(c_2^o - c_2^p - 2)}{2} + u_7 - u_3 > 0.
\end{aligned}$$

We discuss different scenarios based on the value of c_2^p and c_2^o as follows:

1. $c_2^p = 0$

   Under this case, if $c_2^o \geq 1$, $(0,0)$ is a KKT point. In fact, it gives the scenario where there is no peering link between two ISPs and each ISP can support its own customer demand by itself.

2. $c_2^p \neq 0$

   Under this case, the condition for $(0,0)$ to be a KKT point is $c_2^o \geq 1$ and $c_2^o - c_2^p \geq 2$.

As a summary, also considering the condition related with y_2, the condition for $(0,0)$ to be a KKT point is as follows:

$$c_i^o \geq 1 \quad \text{and} \quad (c_i^o - c_i^p \geq 2 \quad \text{or} \quad c_i^p = 0) \quad (i = 1, 2).$$

In fact, $c_i^o \geq 1 (i = 1, 2)$ can be considered as *feasibility condition*, since it guarantees feasibility, namely each ISP can independently satisfy its subscriber demand respectively. The remaining condition means either there is no peering link, or the peering link has significantly lower capacity than the private links such that there is no local traffic exchange under optimal routing.

The above analysis can also be illustrated by Figure A.1. In order for $(0,0)$ to be the minimum, $(y_1^{'}, y_2^{'})$ must be in the third quadrant and $(0,0)$ must be in the feasible region. The feasibility condition gives the requirement $c_i^o \geq 1$, and the negative condition gives the requirement $c_i^o - c_i^p \geq 2$ or $c_i^p = 0$.

A.2 Unlimited Capacity Peering: $u_9 \neq 0$

Let us assume that the minimum point satisfies $u_9 \neq 0$. Therefore we must have $g_9 = y_1 + y_2 - 1 = 0$ which implies $y_2 = 1 - y_1$. We can substitute $1 - y_1$ for y_2 and rewrite the optimization

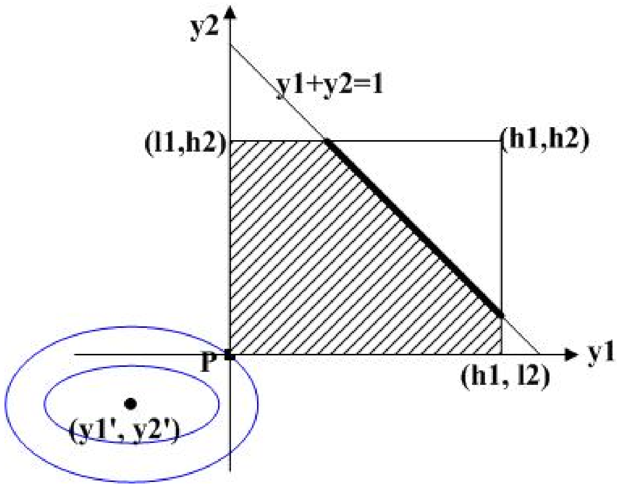

Figure A.1: Conditions for No Peering Scenario

problem as follows:

$$\begin{array}{rll} \mathbf{min}_{y_1} & f(y_1) = y_1^2 + (2\alpha_1(y_2^{'} - 1) - 2\alpha_2 y_1^{'})y_1 + W & \\ \textbf{subject to} & g_1 = -y_1 \leq 0 \qquad\qquad g_2 = y_1 - 1 \leq 0 & \\ & g_3 = 1 - c_2^o - y_1 \leq 0 \quad g_4 = y_1 - c_1^o \leq 0 & \text{(A.5)} \\ & g_7 = y_1 - c_2^p \leq 0 \qquad\quad g_8 = 1 - y_1 - c_1^p \leq 0 & \\ \textbf{where} & y_1^{'} = \dfrac{c_2^p - c_2^o + 2}{4}, \qquad y_2^{'} = \dfrac{c_1^p - c_1^o + 2}{4}. & \end{array}$$

In order to simply the equation, we use W to represent all terms which is not related to y_1 and only depends on c_i^o and c_i^p. After simplification, we still have six constraints, and thus $2^6 = 64$ cases, which makes the analysis very complicated. So we will use Figure A.2 and Figure A.3 to illustrate the condition and solution for this scenario instead of applying the traditional KKT method.

A.2.1 Condition for Unlimited Capacity Peering

As shown in Figure A.2, there must be two intersection points between the feasible rectangle and the line $y_1 + y_2 = 1$ since $l_1 \geq 0$ and $h_1 \leq 1$. Let y_{1l} and y_{1h} denote the x-coordinations

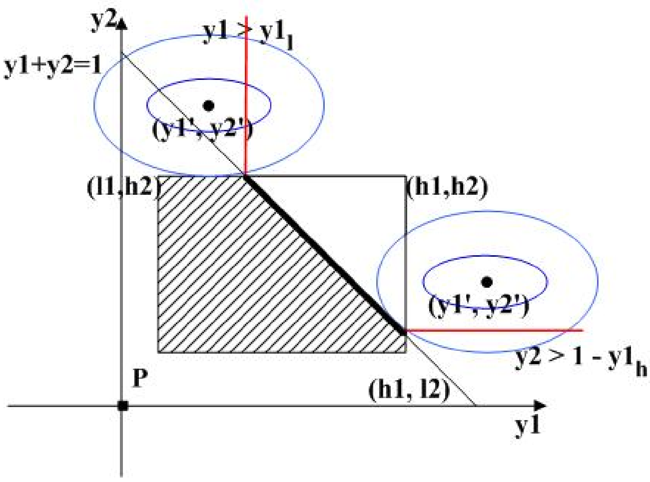

Figure A.2: Unlimited Capacity Peering Scenario: Case 1.

of the two intersection points and $y_{1l} < y_{1h}$. The part between these two intersection points of the line $y_1 + y_2 = 1$, which is the bold line in the figure, gives all points that satisfy both the feasibility conditions and $y_1 + y_2 = 1$. Let us call the bold part of the line *unlimited peering solution line*. Because the objective function is convex, in Figure A.2, we can make the following observation: In order to make the minimum appear on the unlimited peering solution line, $(y_1^{'}, y_2^{'})$ must be in the area that is bounded by three lines: $y_1 + y_2 \geq 1$, $y_1^{'} \geq y_{1l}$ and $y_2^{'} \geq 1 - y_{1h}$.

In Figure A.2, the intersection points appear on the top and the right boundary of the feasible rectangle, which is not always true. Figure A.3 shows the case where the intersection points are on the left and the bottom boundary of the feasible rectangle. In this case, the condition for the minimum to appear on the unlimited peering solution line is $y_1^{'} \geq y_{1h}$ or $y_2^{'} \geq 1 - y_{1l}$ or $y_1^{'} + y_2^{'} \geq 1$.

Let us analyze the condition for unlimited capacity peering scenario as follows:

1. if $c_2^o > c_1^p$ and $c_1^p < 1$ (derived from $1 - max(0, 1 - c_2^o) =$

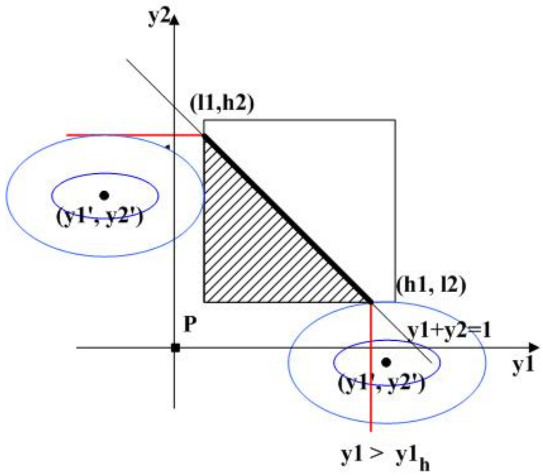

Figure A.3: Unlimited Capacity Peering Scenario: Case 2.

$min(1, c_2^o) > min(1, c_1^p)$), the left intersection point lies on the top boundary, and we have $y_{1l} = max(0, 1 - c_1^p)$, and the condition for unlimited peering scenario related to this intersection point, which is represented by the red line in the figure, is $y_1^{'} \geq y_{1l}$. Otherwise, if $c_2^o \leq c_1^p$ or $c_1^p \geq 1$, the intersection point lies on the left boundary of the rectangle[1], and we have $y_{1l} = max(0, 1 - c_2^o)$, and the condition related to this intersection point for unlimited peering scenario is $y_2^{'} \geq 1 - y_{1l} = min(1, c_2^o)$.

2. if $c_1^o > c_2^p$ and $c_2^p < 1$ (derived from $1 - max(0, 1 - c_1^o) = min(1, c_1^o) \geq min(1, c_2^p)$), the right intersection point lies on the right boundary, and we have $y_{1h} = min(1, c_2^p)$, and the condition related to this intersection point for unlimited peering scenario is $y_2^{'} \geq max(0, 1 - c_2^p)$. Otherwise, if $c_1^o \leq c_2^p$ or $c_2^p \geq 1$, the intersection point lies on the bottom boundary of the rectangle[2], and we have $y_{1h} = min(1, c_1^o)$, and the condition related to this intersection point for unlimited peering scenario is $y_1^{'} > y_{1h} = min(1, c_1^o)$.

[1]Also includes the left top corner.
[2]Also includes the right bottom corner.

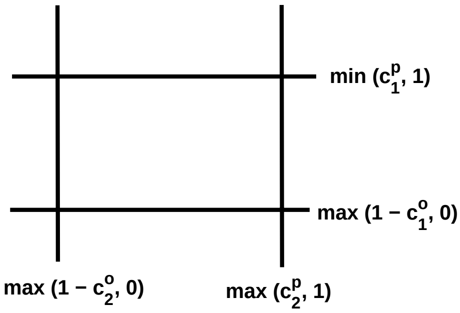

Figure A.4: Location of Intersection Points

Definition 1 *If $c_i^o > c_j^p$ and $c_j^p < 1$, we say ISP_i is selfish provisioning. Otherwise, $c_i^o \leq c_j^p$ or $c_j^p \geq 1$, we say ISP_i is altruistic provisioning.*

Based on the above definitions, we have:

1. both ISPs are selfish provisioning

   In this case, the condition for unlimited peering to occur is $y_1^{'} + y_2^{'} \geq 1$, $y_1^{'} \geq y_{1l}$ and $y_2^{'} \geq 1 - y_{1h}$, which implies

   $$c_1^p + c_2^p - c_1^o - c_2^o \geq 0 \quad \text{and} \quad c_j^o \leq 4c_i^p + c_j^p - 2 \;\; (i, j = 1, 2).$$

2. ISP_1 is selfish provisioning, and ISP_2 is altruistic provisioning

   In this case, $y_{1l} = max(0, 1 - c_2^o)$ and $y_{1h} = min(1, c_2^p)$. The condition for unlimited peering scenario to occur is:

   $$\begin{array}{ll} & c_1^p - c_1^o \geq 4\; min(c_2^o, 1) - 2; \\ \text{or} & (c_1^o \leq 4c_2^p + c_1^p - 2 \;\text{ and }\; c_1^p + c_2^p - c_1^o - c_2^o \geq 0). \end{array}$$

3. ISP_2 is altruistic provisioning, and ISP_1 is selfish provisioning

   This case is symmetrical with case 2.

4. both ISPs are altruistic provisioning

   In this case, $y_{1l} = max(0, 1-c_2^o)$ and $y_{1h} = min(1, c_1^o)$. The condition for unlimited peering scenario to occur is:

   $$c_j^p - c_j^o \geq 2 \quad \text{or} \quad c_j^p - c_j^o + 2 \geq 4c_i \quad \text{or} \quad c_1^p + c_2^p - c_1^o - c_2^o \geq 0.$$

A.2.2 Solution for Unlimited Capacity Peering

In the last subsection, we discussed the conditions for unlimited capacity peering to occur. If these conditions are satisfied, the solution should lie on the unlimited peering solution line, (*i.e.* the part between two intersection point of the line $y_1 + y_2 = 1$), which means $y_{1l} \leq y_1^* \leq y_{1h}$. So in unlimited capacity peering scenario, the routing optimization problem is equivalent to the following one:

$$\begin{array}{rl} \mathbf{min}_{y_1} & (y_1 - \frac{1}{2} - \frac{\alpha_2(c_2^p - c_2^o) - \alpha_1(c_1^p - c_1^o)}{4})^2, \\ \textbf{subject to} & y_1 \in [y_{1l}, y_{1h}], \end{array}$$

where y_{1l} and y_{1h} are x-coordinates of two intersection points of the line $y_1 + y_2 = 1$ and the feasible rectangle. In this simplified form, the optimal solution can be easily derived:

$$y_1^* = \begin{cases} \frac{1}{2} + t & : \text{ if } t \in [y_{1l}, y_{1h}] \\ y_{1l} & : \text{ if } t < y_{1l} \\ y_{1h} & : \text{ if } t > y_{1h} \end{cases}$$

where $t = \dfrac{\alpha_2(c_2^p - c_2^o) - \alpha_1(c_1^p - c_1^o)}{4}$. The break-even prices of two ISPs are [3]:

$$p_1^*(\infty) = \frac{q}{n\alpha_1}(1 - y_2^*) = \frac{q}{n\alpha_1}(\frac{1}{2} + t),$$
$$p_2^*(\infty) = \frac{q}{n\alpha_2}(1 - y_1^*) = \frac{q}{n\alpha_2}(\frac{1}{2} - t).$$

[3] We ignore the boundary cases here.

A.3 Limited Capacity Peering: $u_9 = 0$

First, let us define this new optimization problem:

$$\begin{array}{ll} \mathbf{min}_{y_1,y_2} & \alpha_2(y_1 - y_1^{'})^2 + \alpha_1(y_2 - y_2^{'})^2, \\ \mathbf{subject\ to} & y_i \in [l_i, h_i]. \end{array} \tag{A.6}$$

Let (y_{1n}, y_{2n}) denote the solution of the optimization problem in Equation A.6. We can prove the following lemma:

Lemma 1 *if $y_1^* + y_2^* < 1$, we have $(y_1^*, y_2^*) = (y_{1n}, y_{2n})$.*

Proof: If (y_{1n}, y_{2n}) satisfies the condition $y_1 + y_2 \leq 1$, obviously we have $(y_1^*, y_2^*) = (y_{1n}, y_{2n})$. Otherwise, $y_{1n} + y_{2n} > 1$, since the objective function is a convex function, the solution of the original routing optimization problem must lie on the line $y_1 + y_2 = 1$, so that $y_1^* + y_2^* = 1$. It contradicts with our assumption in the lemma. Therefore we must have $(y_1^*, y_2^*) = (y_{1n}, y_{2n})$ in the limited capacity peering scenario. ∎

Based on this lemma, we can solve the optimization problem in Equation A.6, and get the routing solution for limited capacity peering scenario as follows:

$$y_i^* = \begin{cases} y_i^{'} & : \text{ if } y_i^{'} \in [l_i, h_i] \\ l_i & : \text{ if } y_i^{'} < l_i \\ h_i & : \text{ if } y_i^{'} > h_i. \end{cases}$$

The break-even prices of two ISPs are:

$$p_1^* = \frac{q}{n\alpha_1}(1 - y_2^*) = \frac{q}{n\alpha_1}(\frac{2 - c_1^p + c_1^o}{4}),$$
$$p_2^* = \frac{q}{n\alpha_2}(1 - y_1^*) = \frac{q}{n\alpha_2}(\frac{2 - c_2^p + c_2^o}{4}).$$

Bibliography

[1] Cernet ipv4. http://bgpview.6test.edu.cn/bgpview/cur_ana/ipv4/ipv4_bgp.shtml.

[2] Wikipedia. http://en.wikipedia.org/wiki/Internet.

[3] Andre Broido, Evi Nemeth and kc claffy. Internet Expansion, Refinement and Churn. *European Transactions on Telecommunications*, 13(1):33–51, Jan-Feb 2002.

[4] S. Banerjee, C. Kommareddy, and B. Bhattacharjee. Efficient peer location on the internet. *Computer Networks*, 45(1):5–17, 2004.

[5] Bradley Huffaker, Marina Fomenkov, Daniel J. Plummer, David Moore and K. Claffy. Distance Metrics in the Internet. In *IEEE International Telecommunications Symposium*, Sept. 2002.

[6] Bruno Quoitin, Cristel Pelsser, Olivier Bonaventure and Steve Uhlig. A performance evaluation of bgp-based traffic engineering. *International Journal of Network Management*, 15(3):177–191, 2005.

[7] Bruno Quoitin, Sebastien Tandel, Steve Uhlig and Olivier Bonaventure. Interdomain Traffic Engineering with redistribution communities. *Computer Communications*, 27(4):355–363, 2004.

[8] Bruno Quoitin, Steve Uhlig, Cristel Pelsser, L. Swinnen and Olivier Bonaventure. Interdomain Traffic Engineering with BGP. In *IEEE Communications Magazine*, 2003.

[9] Byung-Gon Chun, Kamalika Chaudhuri, Hoeteck Wee, Marco Barreno, Christos H. Papadimitriou and John Kubiatowicz. Selfish Caching in Distributed Systems: A Game-theoretic Alysis. In *PODC '04: Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, pages 21–30, New York, NY, USA, 2004. ACM Press.

[10] Chris Metz. Interconnecting ISP Networks. *Internet Computing, IEEE*, 5(2), 2001.

[11] David K. Goldenberg, Lili Qiu, Haiyong Xie, Yang Richard Yang and Yin Zhang. Optimizing cost and performance for multihoming. In *SIGCOMM '04: Proceedings of the 2004 conference on applications, technologies, architectures, and protocols for computer communications*, pages 79–92, New York, NY, USA, 2004. ACM Press.

[12] Feng Wang and Lixin Gao. On Inferring and Characterizing Internet Routing Policies. In *IMC '03: Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 15–26, New York, NY, USA, 2003. ACM Press.

[13] Giuseppe Di Battista, Maurizio Patrignani, Maurizio Pizzonia and Massimo Rimondini. Towards Optimal Prepending for Incoming Traffic Engineering. In *3rd International Workshop on Internet Performance, Simulation, Monitoring, and Measurement (IPS MoMe 2005)*, 2005.

[14] Hyunseok Chang and Sugih Jamin and Morley Mao and Walter Willinger. An Empirical Approach to Modeling

Inter-AS Traffic Matrices. In *IMC '05, 2005 Internet Measurement Conference*, Berkley, CA, USA, 2005.

[15] I. Beijnum. *BGP*. O'Reilly, 2002.

[16] J Glasner. P2P Fuels Global Bandwidth Binge. *Wired News*, April 2005.

[17] J. V. Outrata. Optimality Conditions for a Class of Mathematical Programs with Equilibrium Constraints. *Mathematics of Operations Research*, 24(3):627–644, 1999.

[18] Jessie Hui Wang, Rocky Chang, D.M. Chiu and John C.S. Lui. Characterizing the Performance and Stability Issues of the AS Path Prepending Method: Taxonomy, Measurement Study and Analysis. In *Proc. ACM SIGCOMM Asia Workshop*, April 2005.

[19] John W. Stewart. *BGP4 : Inter-domain Routing in the Internet*. Addison Wesley, 1999.

[20] Jong-Shi Pang and Masao Fukushima. Quasi-variational Inequalities, Generalized Nash Equilibria, and Multi-leader-follower Games. *Computational Management Science*, 2(1):21–56, Jan 2005.

[21] Kannan Varadhan, Ramesh Govindan and Deborah Estrin. Persistent Route Oscillations in Inter-domain Routing. Technical Report ISI 96-631, Univ. of Southern California Information Sciences Institute, Marina del Rey, CA, 1996.

[22] Kannan Varadhan, Ramesh Govindan and Deborah Estrin. Persistent Route Oscillations in Inter-domain Routing. *Computer Networks*, 32(1):1–16, January 2000.

[23] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy and John Zahorjan. Mea-

surement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles*, Bolton Landing, New York, October 2003.

[24] Kurt Tutschku and Phuoc Tran-Gia. Traffic Characteristics and Performance Evaluation of Peer-to-Peer Systems. In *Peer-to-Peer Systems and Applications*, pages 383–397. Springer, 2005.

[25] L. Swinnen, S. Tandel, S. Uhlig, B. Quoitin and O. Bonaventure. An evaluation of bgp-based traffic engineering techniques. www.info.ucl.ac.be/people/OBO/papers/cost263-chapter.pdf.

[26] Lixin Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, 2001.

[27] Lixin Gao and Jennifer Rexford. Stable Internet Routing without Global Coordination. In *Measurement and Modeling of Computer Systems*, pages 307–317, 2000.

[28] Martin B. Weiss and Seung Jae Shin. Internet Interconnection Economic Model and its Analysis: Peering and Settlement. *Netnomics*, 6(1):43–57, 2004. available at http://ideas.repec.org/a/kap/netnom/v6y2004i1p43-57.html.

[29] Masao Fukushima and Jong-Shi Pang. Some Feasibility Issues in Mathematical Programs with Equilibrium Constraints. *SIAM Journal of Optimization*, 8:673–681, 1998.

[30] Nick Feamster, Jared Winick and Jennifer Rexford. A Model of BGP Routing for Network Engineering. In *SIGMETRICS '04/Performance '04: Proceedings of the joint*

international conference on Measurement and modeling of computer systems, pages 331–342, New York, NY, USA, 2004. ACM Press.

[31] Nick Feamster, Jay Borkenhagen, and Jennifer Rexford. Controlling the Impact of BGP Policy Changes on IP Traffic. Technical Report 011106-02, AT&T Research, November 2001.

[32] Nick Feamster, Jay Borkenhagen and Jennifer Rexford. Guidelines for Interdomain Traffic Engineering. *SIGCOMM Computor Communnications Review*, 33(5):19–30, 2003.

[33] Raj Jain, Dah Ming Chiu and W. Hawe. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. DEC Research Report TR-301, Digital Equipment Corporation, Maynard, MA, USA, Sep 1984.

[34] Ram Keralapura, Nina Taft, Chen-Nee Chuah and Gianluca Iannaccone. Can ISPs Take the Heat from Overlay Networks? In *ACM HotNets III*, November 2004.

[35] Ramesh Govindan, Cengiz Alaettinoglu, George Eddy, David Kessens, Satish Kumar, Wee-San Lee. An Architecture for Stable, Analyzable Internet Routing. *Network, IEEE*, 13(1):29–35, Jan/Feb 1999.

[36] Ramesh Johari and John Tsitsiklis. Routing and Peering in a Competitive Internet. MIT, Tech., 2003.

[37] Richard Bellman. On a Routing Problem. *Quarterly of Applied Mathematics*, 16(1):87–90, 1958.

[38] Rocky Chang and Michael Lo. Inbound Traffic Engineering for Multihomed ASes using AS Path Prepending. *IEEE Network*, 19(2), March 2005.

[39] Route Views Project. http://www.routeviews.org/.

[40] Ruchir Bindal, Pei Cao, William Chan, Jan Medval, George Suwala, Tony Bates and Amy Zhang. Improving Traffic Locality in BitTorrent via Biased Neighbor Selection. In *Distributed Computing Systems, 2006. ICDCS 2006. 26th IEEE International Conference on*, pages 66–77, 2006.

[41] Ruomei Gao and Constantinos Dovrolis and Ellen W. Zegura. Interdomain Ingress Traffic Engineering Through Optimized AS-Path Prepending. In *NETWORKING*, pages 647–658, 2005.

[42] S. Leyffer and T. Munson. Solving Multi-Leader-Follower Games. *Preprint ANL/MCS-P1243-0405*, 2005.

[43] Samantha Lo, Rocky K. C. Chang and Lorenzo Colitti. An Active Approach to Measuring Routing Dynamics Induced by Autonomous Systems. In *Proceedings of ACM Workshop on Experimental Computer Science*, June 2007.

[44] Srinivas Shakkottai and R. Srikant. Economics of Network Pricing with Multiple ISPs. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, pages 184–194 vol. 1, 2005.

[45] Srinivasan Seetharaman and Mostafa Ammar. On the Interaction between Dynamic Routing in the Native and Overlay Layers. In *INFOCOM 2006*, Barcelona, Spain, April 2006.

[46] Steve Uhlig and Olivier Bonaventure. Implications of interdomain traffic characteristics on traffic engineering. *European Transactions on Telecommunications*, January 2002.

[47] Subhabrata Sen and Jia Wang. Analyzing Peer-to-peer Traffic Across Large Networks. *IEEE/ACM Transaction on Networks*, 12(2):219–232, 2004.

[48] Timothy G. Griffin and Gordon T. Wilfong. A Safe Path Vector Protocol. In *Proceedings of INFOCOM*, volume 2, pages 490–499, 2000.

[49] Timothy G. Griffin and Gordon Wilfong. An Analysis of BGP Convergence Properties. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 277–288, New York, NY, USA, 1999. ACM Press.

[50] Timothy G. Griffin, F. Bruce Shepherd and Gordon Wilfong. Policy Disputes in Path-Vector Protocols. In *Proc. Seventh International Conference on Network Protocols, 1999 (ICNP '99)*, pages 21–30, Nov. 1999.

[51] Timothy G. Griffin, F. Bruce Shepherd and Gordon Wilfong. The Stable Paths Problem and Interdomain Routing. *IEEE/ACM Transactions on Networking*, 10(2):232–243, 2002.

[52] Timothy Griffin and Gordon Wilfong. Analysis of the MED Oscillation Problem in BGP. Paris, France, November 2002. 10th IEEE International Conference on Network Protocols (ICNP'02).

[53] Wenjie Jiang, Dah-Ming Chiu and John C. S. Lui. On the Interaction of Multiple Overlay Routing. *Performance Evaluation*, 62(1-4):229–246, 2005.

[54] William B. Norton. Interconnection Strategies for ISPs. Technical report, Equinix, 1999.

[55] William B. Norton. Internet Service Providers and Peering. Technical report, Equinix, 2001.

[56] William B. Norton. The Art of Peering: The Peering Playbook. Technical report, Equinix, 2001.

[57] William B. Norton. A Business Case for ISP Peering. Technical report, Equinix, 2002.

[58] William B. Norton. The Evolution of the US Internet Peering Ecosystem. Technical report, Equinix, 2003.

[59] Yin Zhang and Matthew Roughan and Nick Duffield and Albert Greenberg. Fast accurate computation of large-scale ip traffic matrices from link loads.

[60] Yong Liu, Honggang Zhang, Weibo Gongy and Don Towsle. On the Interaction between Overlay Routing and Underlay Routing. In *Proceedings of 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 4, pages 2543– 2553, March.

[61] Z. Morley Mao. BGP Beacon Info. http://psg.com/ zmao/ BGPBeacon.html.