CITY UNIVERSITY OF HONG KONG
香港城市大學

A Study of Network Neutrality and Differentiated Services
網絡中立性與差異化服務的研究

Submitted to
Department of Electrical Engineering
電機工程系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

Li Fan
李帆

December 2020
二零二零年十二月

Abstract

During the last two decades, we have witnessed a continuous debate on whether net neutrality policies should be adopted to maintain the openness of the Internet. This debate has led the U.S. Federal Communications Commission to adjust its Internet regulation policies several times in the U.S., while there is still a lack of understanding of net neutrality-related issues among the public. Network regulation policies are important because society increasingly relies on the Internet. With our ever-increasing dependence and use of the Internet, such policies have a significant impact.

This multi-disciplinary study aims to provide useful information for analyzing and understanding the network neutrality debate. It begins with a comprehensive literature study on net neutrality from three perspectives: policy, economic, and engineering. From this literature study, we found that sometimes proponents and opponents of net neutrality policy are not clearly distinguished: A party considered to be a proponent of net neutrality may also adopt actions that in fact violate its policy.

Aiming to achieve the two opposing objectives, namely, to follow the spirit of network neutrality, and to allow differentiated services, this thesis uses queuing models to comparatively study the tradeoffs between the economic cost of ISPs, and the delay of end-users. The results demonstrate that offering all users the same bandwidth is not an optimized solution to meet the requirements of different services. In addition, we differentiate data packets by their transmission protocols, such as TCP and UDP, to examine the quality of services perceived by users. By using real datasets, we demonstrate that this approach can improve

users' experience while using the same or even less bandwidth than treating all packets in the first come first served manner.

Alongside the performance analysis on net neutrality, we conducted a public opinion survey on network regulation policies in Hong Kong. We emphasize two of the key issues about net neutrality, namely zero-rating and differentiated prioritized services in Internet access. Since the concept of net neutrality is rather elusive and complex, we use novel 2×2 factorial vignettes to illustrate the network regulation policies to be implemented in Hong Kong. The polling data show that most of the respondents acknowledged the importance of the network neutrality policy for Hong Kong while also accepting both zero-rating and differentiated services.

CITY UNIVERSITY OF HONG KONG
Qualifying Panel and Examination Panel

Surname:	LI
First Name:	Fan
Degree:	PhD
College/Department:	Department of Electrical Engineering

The Qualifying Panel of the above student is composed of:

Supervisor(s)

Prof. ZUKERMAN Moshe	Department of Electrical Engineering City University of Hong Kong

Co-Supervisor(s)

Prof. Richard M WALKER	Department of Public Policy City University of Hong Kong

Qualifying Panel Member(s)

Dr. CHAN Chi Hung Sammy	Department of Electrical Engineering City University of Hong Kong
Dr. WONG Wing Ming Eric	Department of Electrical Engineering City University of Hong Kong

This thesis has been examined and approved by the following examiners:

Dr. PO Lai Man	Department of Electrical Engineering City University of Hong Kong
Dr. YEUNG Kai Hau Alan	Department of Electrical Engineering City University of Hong Kong
Prof. ZUKERMAN Moshe	Department of Electrical Engineering City University of Hong Kong
Prof. FONSECA Nelson	Computer Systems Department Institute of Computing of The State University of Campinas

I would like to dedicate this thesis to my parents.

Acknowledgements

I have benefited significantly from interaction with many people at the City University of Hong Kong, I would like to thank them.

Foremost, I would like to express my sincere thanks to my supervisor, Prof. Moshe Zukerman, Department of Electronic Engineering, for his kindness, guidance, and support before and during my PhD studies.

My sincere thanks also go to Prof. Richard Walker, Department of Public Policy, City University of Hong Kong. Thanks to his project and ideas that have made this PhD work. His knowledge, experience, and skills in public polling have supported my PhD work much. Meanwhile, Dr. Lee Myoung-Jin taught me the knowledge of data collecting and relevant skills of data statistics.

My gratitude also goes to my former colleagues Dr. Ji Li, Dr. Shuo Li, Dr. Jiongze Chen, Dr. Cong Cao, Dr. Zengfu Wang, Dr. Jingjin Wu, Dr. Anliang Cai, Dr. Chang Xing, Dr. Qing Wang, Mr. Wang Xinyu, Miss Wang Tianjiao, and Miss Xu Jiahe who in the past five years gave me time to discuss problems.

As the staff gave me strong support during my PhD study, I wish to express my thanks to Ms. Ng Shirley, who helped me with setting up an experimental platform. I am grateful to Ms. Lucia Mak and Ms. Mimi Lee, who helped me handle various administrative work. Thanks also go to Ms. Charis So from the Human Resource Office of CityU.

Finally, I wish to dedicate this thesis to my parents for their endless love and support.

The work described in this thesis and the related expenditure of the questionnaire survey were supported in part by the grant from the project entitled “Analysis of Net Neutrality from a Technical, Economic and Policy Perspectives” [CityU Project No. 9610638], in part by the City/KTO research project 6354025.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

ANOVA Analysis of variance

ARPANET Advanced Research Projects Agency Network

AS Autonomous System

BBS Bulletin board system

CAP Content Application Provider

CDN Content Distribution Network

CP Content Provider

CSP Content Service Provider

DCF Distributed Coordination Function

DiffServ Differentiated Services

DSL Digital Subscriber Line

EVM Experiment vignette method

FCC Federal Communication Commission

FTC Federal Trade Commission

FTTH Fiber To The Home

GoS Grade of Service

HTTP Hyper Text Transfer Protocol

ICMP Internet Control Message Protocol

ICT Information and communications technology

IETF Internet Engineering Task Force

IntServ Integrated Service

IP Internet Protocol

ISO International Organization for Standardization

ISP Internet Service Provider

LRD Long Range Dependent

MPLS Multi-Protocol Label Switching

OSI Open System Interconnection

OSPF Open Shortest Path First

PMP Paris Metro Pricing

PPBP Poisson Pareto burst process

PSTN Public Switched Telephone Network

QoE Quality of Experience

QoS Quality of Service

SLA Service Level Agreement

SMTP Simple Message Transfer Protocol

TCP Transmission Control Protocol

ToS Type of Service

UDP User Datagram Protocol

VoD Video On Demand

VoIP Voice over Internet Protocol

VPN Virtual Private Network

Chapter 1

Introduction

1.1 Motivation

The Internet has become an essential element of today's life. It has penetrated into nearly all aspects of peoples' life since it is the biggest information carrier in the world. Therefore, it is important to maintain its normal operations.

The Internet has been operated under a non-discriminating rule since its inception. Most of the information is carried by the Internet under the "first come first serve " manner in the form of data packets that are served by the TCP/IP (Transmission Control Protocol / Internet Protocol) protocol suites with the "Best Effort" service. This means that the network does not guarantee the successful transmission of data, nor does it guarantee the user's service quality or a certain priority. In its early stage, Internet services were not as varied as they are now. The majority of the information was message and images, which were mainly the content of emails and Bulletin board system (BBS). Although the bandwidth resources were not as abundant as they are today, the transmission delay of text messages could be tolerated by the users. As technology developed, multimedia service rapidly penetrated into the Internet. Voice calls can be made, and video clips are increasingly prevalent, and media-rich web pages keep on growing. The average data volume of a web page doubled from 2010 to 2013

[9], and video is the major consumer of bandwidth today. In 2014, video content represented 64% of the total Internet traffic, and it is estimated that videos will take up 80% of the total Internet traffic volume [10]. In contrast with the delay of text and image, the voice and videos' delay can be much more easily perceived by human beings, therefore the tolerance of voice and videos' delay is much smaller, especially the online voice communication and video broadcasting. Different users have different needs, given this situation, the tiered service class of data communications is in demand.

Certain Internet services create conflicts of interest for ISPs (Internet service providers). A typical example is VoIP (Voice over Internet Protocol) which impacts the telephone market. This has led ISPs to differentiated Internet traffic through technical means, they banned or throttled the services which conflicted with existing services. However, not all Internet services cause loss of ISPs' interest. A few CPs (content provider) pay extra money to ISPs to let their content have higher priority than other traffic in order to provide a better quality of service to end-users.

The existence of differentiated traffic on the Internet aroused the attention of somebody who wanted to make all the traffic to be treated equally. After the term "Net Neutrality" was formally coined [11] by Tim Wu in 2003, network neutrality has been hotly debated. Net neutrality is also sometimes called "Internet openness" [12]. In 2015, Federal Communication Commission (FCC) invited the public to vote on "Open Internet", four million respondents voted and most of them favored regulations. As a result, FCC set three bright-line rules for net neutrality [13]: "no blocking, no throttling, and no paid prioritization."

There is no indication that any agreement has been made on network neutrality. It is extensively debated even at the time of writing. In this thesis, we try to provide a comprehensive view of net neutrality, instead of standing at a certain perspective only. Policymakers define net neutrality from different aspects [14, 15], and justify regulation rules to guarantee the Internet to be "fair", economic researchers propose different economic models to describe the Internet's ecosystem, especially discuss the gain and loss of the stakeholders. While the engineers investigate the issue from the angle that maximizes the performance of the communication networks, such as minimum the blocking probability, increasing the link utilization, reduce the probability of occurrence of failure. From a macro perspective, the whole Internet can be divided into three tiers. The issues of differentiating Internet traffic mainly locates in the access (Tier 3) networks, where the ISPs connect to, as shown in Fig. 1.1. Outside the Tier 3 network, the network management such as the core network (Tier 1) is running autonomously. Therefore, the scope of this work is mainly focused on the Tier 3 network.

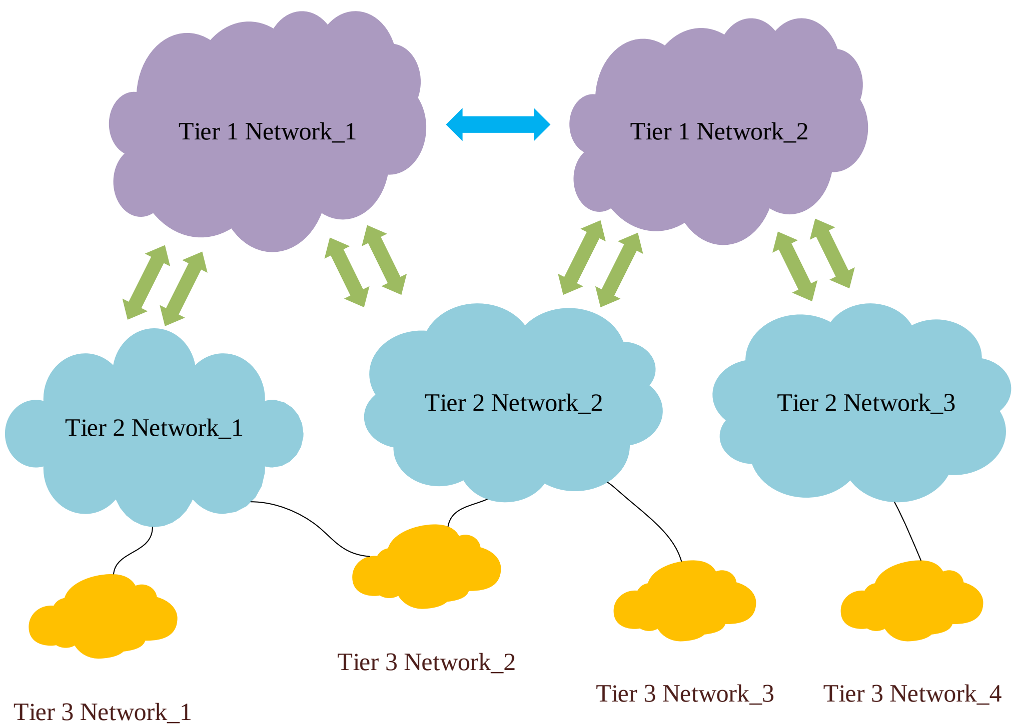


Fig. 1.1 The tier structure of the Internet

1.2 Organization

This work focuses on the analysis and modeling of network neutrality. The outline of this thesis is as follows.

Given the complexity of network operation and management policy, Chapter 2 introduces the debate on net neutrality and gives a comprehensive study of network management from policy, economic and engineering perspectives.

In Chapter 3, we review the technical standards of differentiated services implementation in the network from a layered model perspective.

In Chapter 4, we study the impact that introducing differentiated services into broadband networks. We also propose a solution to address the dispute of network neutrality by differentiating data packets by service protocols and test this proposal with real data sets.

In Chapter 5, we carry out a public survey on network management policies and analyze the results of the survey.

Conclusion and future work are given in Chapter 6.

1.3 Contributions

Net neutrality has attracted increasing attention and it is being extensively debated at the time of the writing. Although there already exists a substantial literature on the topic of net neutrality, most of the literature only focuses on single perspectives such as policy-making or economic considerations. In addition to legal and policy perspectives, the discussions in this thesis also include economic and engineering perspectives. This work does not try to throw a clear-cut conclusion on whether or not a network regulation policy should be adopted, but provides the reader with a better understanding of the concept of net neutrality. It needs to be made clear that some regulatory policies have the opposite effect, and some net neutrality advocates are not neutral, but anti-competitive. In contrast, actions that seem to undermine net neutrality have improved social welfare. We believe such research work is helpful for establishing regulation of network management while balancing equity and efficiency.

The main contributions of this thesis are fourfold:

1. A comprehensive survey on net neutrality from perspectives including policy, economic, and technology is presented. This comprehensive review tries to help people understand what is "net neutrality".

2. The key issue of "differentiated services (or paid prioritized services) provided by ISPs" is simulated in both wireline and wireless networks, benefits for service providers and the impact to end-users are provided.

3. A scheme that prioritizes data packets through transmission protocols is proposed and tested using real datasets. The results show that it can indeed improve the end-users' experience and help to reduce disputes about network neutrality.

4. A public questionnaire survey of net neutrality is conducted. The survey is limited to Hong Kong SAR, where the Internet's access rates ranks are one of the highest in the Asia pacific. We believe such a survey has practical implications for policymaking in Hong Kong.

1.4 Publications

1.4.1 Journal Publications During Author's PhD Study:

- C. Xing, R. G. Addie, Y. Peng, R. Lin, **F. Li**, W. Hu, V. Abramov and M. Zukerman, "Resource Provisioning for a Multi-layered Network" *IEEE Access*, vol. 7, no. 1, pp. 16226-16245, 2019.

- Q. Xu, **F. Li**, J. Sun and M. Zukerman, "A New TCP/AQM System Analysis" *Journal of Network and Computer Applications*, vol. 57, pp. 43-60, November 2015.

1.4.2 Conference Publications During Author's PhD Study:

- **F. Li**, M. Zukerman, M. Jin Lee, R. Cheung and Richard M. Walker, "Public Policy Polling on Smart City: Autonomous Vehicles and Net Neutrality," in *The 4th Annual Conference of Asia-Pacific Public Policy Network*, Hong Kong, March 2019.

- Y. Peng, R. Lin, **F. Li**, C. Xing, J. Guo, W. Hu, V. Abramov, R. G. Addie and M. Zukerman, "Validation of multi-layer network optimization," in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, Italy, August 2016.

- M. Wang, L. Hou, E. W. M. Wong, J. Guo, J. Liu, **F. Li**, A. Cai, H. Mehrvar, D. Wang, D. Geng, E. Bernier, and M. Zukerman, "Design implications of the add/drop ratio in transparent photonic networks," in *2015 17th International Conference on Transparent Optical Networks (ICTON)*, Budapest, Hungary, July 2015.

1.4.3 Early Publications By Author:

- **F. Li**, J. Sun, M. Zukerman, Z. Liu, Q. Xu, S. Chan, G. Chen, and K. T. Ko, "A comparative simulation study of TCP/AQM systems for evaluating the potential of neuron-based AQM schemes," *Journal of Network and Computer Applications*, vol. 41, pp. 274-299, May, 2014.

- Z. Rosberg, J. Li, **F. Li**, and M Zukerman, "Flow scheduling in optical flow switched (OFS) networks under transient conditions," *Journal of Lightwave Technology*, vol. 29 no. 21, pp. 3250-3264, September 2011.

- F. Ge, S. Chan, L. L. H. Andrew, **F. Li**, L. Tan, and M. Zukerman, "Performance effects of two-way FAST TCP," *Computer Networks*, vol. 55 no. 13, pp. 2976-2984, September 2011.

- V. Abramov, R. G. Addie, **F. Li**, Y. Peng, and M. Zukerman, “A new teletraffic approach for network planning and evolution prediction,” in *Proc. OFC/NFOEC 2012*, Los Angeles, March 2012.

- R. G. Addie, D. Fatseas, Y. Peng, **F. Li**, and M. Zukerman, “How good (or bad) is shortest path routing in layered networks,” in *Proc. ATNAC 2012, Brisbane*, Australia, November 2012.

Chapter 2

An Overview of Net Neutrality

In general, the net neutrality principle stipulates that ISPs should “treat all the data equally, without discrimination”. Net neutrality (sometimes also called “network neutrality”) mainly applies to the last mile, which is the final leg of the network that delivers telecom services to end-users. Although there are many publications related to the topic of net neutrality, most of them only focus on policy and on legal perspectives. In addition to legal and policy issues, this chapter investigates also economic and engineering perspectives.

2.1 The Debate and Iconic Events

The “Net neutrality” debate has been ongoing even before it was conceptually introduced by Tim Wu in 2003 [11, 16]. In February 2004, Michael Powell, who was the FCC chairman at the time, set up four principles for net neutrality policy [17, 18]. These four principles, also called the four “Internet Freedoms”, reflect his expectation that the broadband industry allows users:

- “to access lawful content;
- to use applications;

- to attach personal devices that do not harm the network;
- to obtain service plan information entitlement to competition."

The necessity for network neutrality regulation is based on the assumption that ISPs have incentives to discriminate against unaffiliated providers [19], and network technologies enable ISPs to do so. At the start of the 21st-century ISPs benefited from the emerging new technologies that provided data and VoIP services, where the latter soon became a strong competitor and a threat to the traditional telephone service providers. The following events demonstrated that FCC's "Internet freedoms" (especially the second and fourth) were not well respected by ISPs:

1. In 2004 and 2005, FCC fined Madison River Communications Company US$15,000, for blocking their customers from using VoIP call. Such VoIP was provided by Vonage, an Internet phone service provider, which is a competitor of Madison River Communications [20].
2. In 2005, by using forged packets, Comcast throttled peer-to-peer (P2P) file-sharing applications. Comcast did not stop restricting the P2P sharing protocols such as BitTorrent until FCC forced them to stop [21].
3. Apple was forced by AT&T to stop using Skype which is a VoIP service in their "iPhone" product, from 2007 to 2009 [21].
4. Similar to the case of Madison River Communication case, in 2008 Comcast interrupted and degraded P2P traffic (especially BitTorrent) on its network. For this reason, Comcast was fined by the FCC since such practice violated the "Internet Policy Statement" [22].

With video streaming becoming prevalent in the late 2000s, more and more end-users watched videos and movies through broadband service. In 2009, FCC Chairman Julius

Genachowski appended two additional “Freedoms of the Internet” [18, 23] (The fifth one was considered to be controversial):

5. “A non-discrimination principle whereby ISPs cannot discriminate against particular content and application providers, though they may engage in “reasonable network management.

6. A consumer transparency principle where ISPs must disclose their network management practices to consumers, content and application providers, and the FCC.”

However, the following events still show that ISPs kept disrespecting Internet freedoms rules:

- In 2011, MetroPCS, a U.S. wireless carrier, announced that all the streaming video services were blocked except YouTube over its cellular network [21].

- After Apple introduced its real-time video call App “FaceTime” in 2010, AT&T announced that it would disable such video-calling function on its subscribers’ iPhones unless they upgraded to a more expensive subscription plan in 2012 [21].

- In 2014, customers complained that Netflix and Hulu did not provide the indicated QoS when they watched the streaming video on demand. A news report [24] revealed that there were special deals between content providers (CPs) and ISPs, with CPs paying extra money to let the ISPs send their multimedia content prior to other contents. Such deals violated the fifth “non-discrimination rule.”

- In July 2017, end-users accused Verizon Wireless that videos from YouTube and Netflix were played slower than usual. Verizon explained that the situation was caused by “network testing” and that “reasonable network management practices” were permitted by net neutrality rules [25].

In June 2018, the FCC had repealed the net neutrality regulation which was previously established by the Obama administration. On October 1, 2019, the district of Columbia Court of Appeals repealed most provisions of net neutrality.

From these events, one can observe that before 2010, end-users mainly focused on whether they can use new applications. Big ISPs blocked or throttled certain applications' usage since such applications eroded their markets. After 2010, with increased network capacity enabling video transmission and prosperity of video CPs such as YouTube and Netflix, end-users were concerned more about the QoS and QoE of their online video streaming. Such QoS and QoE requirements give ISPs opportunities and incentives to use differentiated and prioritized services. While the non-discrimination rule may be interpreted as not allowing such practices as differentiation and prioritization, ISPs could use the term "special network management" that appeared in the fifth net neutrality rule as a justification for such practices.

Proponents of network regulations are mainly end-user organizations and CPs, while ISPs stand on the opposite side. The central issue of net neutrality is whether differentiated (or prioritized) services should be allowed at the last mile of Internet access. The network is regarded as non-neutral if it allows differentiated services (with differentiated pricing) to exist.

2.1.1 Definition of "Net Neutrality"

So far, there is no well-recognized and rigorous definition of "Net neutrality" [26–28]. Generally, it means that ISPs only charge customers once for access, and that they are not allowed to charge CP for sending data to customers, and favor one CP over another.

Wu [11] pointed out that: "The basic principle behind a network anti-discrimination regime is to give users the right to use non-harmful network attachments or applications, and give innovators the corresponding freedom to supply them." Lessig [29] pointed out that the innovative development and prosperity of the Internet are credited to its “end-to-end” architectural design. Therefore competition is critical to protect end-to-end neutrality.

Different definitions of Net Neutrality have been proposed during the debate on the issue. A more recent paper [26] summarized a list of definitions from various aspects:

- “Net neutrality prohibits ISPs from speeding up, slowing down or blocking Internet traffic based on its source, ownership or destination (strict neutrality).

- Net neutrality usually means that broadband service providers charge consumers only once for Internet access, do not favor one CP over another, and do not charge CP for sending information over broadband lines to end-users.

- A person engaged in the provision of fixed broadband Internet access service, insofar as such person is so engaged, shall publicly disclose accurate information regarding the network management practices, performance, and commercial terms. The person shall not block lawful content, applications, services, or non-harmful devices, subject to reasonable network management. The person shall not unreasonably discriminate in transmitting lawful network traffic over a consumer’s broadband Internet access service.”

2.1.2 Why Net Neutrality?

Large online companies such as Facebook, Google, Amazon Prime, and Netflix support net neutrality regulations. They believe that the development and prosperity of Internet services in the past decade should be attributed to net neutrality. They may argue that large ISPs may use their power to stifle new startups and new innovations [30], and this may adversely affect the economic growth and prosperity gained from the Internet.

2.1.3 Why Not Net Neutrality?

On the other hand, ISPs such as Comcast, Verizon and AT&T argue that allowing differentiated services is beneficial to the Internet ecosystem. They argue for a non-neutral network for the following reasons: First, prioritized services, as an approach of network management is more efficient to deal with security issues and network congestion. Second, without differentiated services and pricing, ISPs will lose the incentive to invest in resources and technologies that increase the network capacity, and this will eventually impair QoS as available capacity cannot accommodate the ever-increasing bandwidth requirement. Third, the revenue model of the two-side broadband market is unbalanced. ISPs only receive a fixed payment from their subscribers monthly (or yearly), while CPs generate far more revenues from video on demand (VoD) and advertisements.

2.1.4 Summary on Net Neutrality Debate

It is not difficult to find out that: Net neutrality proponents mainly consist of CPs and end-users' groups, who care about the innovation of network and social welfare that the Internet provides; while net neutrality opponents are mainly ISPs, who care about whether they can get revenue from their investment on the network infrastructure.

Many researchers have summarized the net neutrality debate [31–36]. Faulhaber [37] reviewed the evolution of the debate. Ganley in [38] pointed out that using a certain regulatory process to ensure end-users' choice is the best solution. In fact, the debate is never-ending. News related to the debate can be found in the media from time to time, such as FCC's repealing of net neutrality [39] and about hearings on net neutrality [40]. In addition to the debate, work around net neutrality was also taking place among different countries; various consultation work can be found in [41].

In the following, we will discuss net neutrality from three perspectives: Policy, Economic, and Engineering. We note, however, that these perspectives are interrelated. Administration policy affects the ISP pricing, and competition affects policymaking. The ever-evolving telecommunication technology lowers the price of broadband services. Nevertheless, for ease of exposition, we separate the discussion into three perspectives.

2.2 The Policy Perspectives

In this section, we discuss the policy perspective of the net neutrality issue and review related concepts, principles, and practices. Peha *et al.* [42] pointed out the complexity of net neutrality from the regulator's point of view: the regulation is handled by different government sectors, each operating separately. For example, in the USA, FCC and FTC (Federal Trade Commission) treat net neutrality as a regulatory issue, but the Department of Justice sees net neutrality as something related to antitrust. Goth [43] reviewed the suggestions given by the regulator FCC and the opinions from major network operators such as AT&T. He also discussed the implementation of net neutrality regulations in Europe and Asia. Cooper in [44] gave a comprehensive comparison of the net neutrality policies in the UK and the US.

Net neutrality is often confused with net diversity. These two concepts have many intersections such as they both concern the social welfare and fairness issues. When analyzing social welfare in the context of net diversity, the following points should be noticed [45]:

- *The robustness of competition*: A guideline suggests that a lower limit is to keep six similarly sized companies to prevent the competition being harmed from using vertical integration. In the broadband service industry, however, it is still questionable that network diversity will guarantee the competition to mitigate the risk of vertical integration.

- *The Heterogeneity of demand*: The assumption that customers' demand is heterogeneous provides the reason for network diversity. In monopolistic competition, the assumption is further expanded that consumer preferences are symmetric with respect to each of the competing groups.

- *The Multi-dimensionality of welfare under network diversity*: When suppliers offer consumers more than one class of commodity/service, they not only offer competitive prices but also better meet particular customers' preferences. In broadband network service, ISPs could offer differentiated services such as fast lane. In mobile network services, ISPs could offer package plans with different data volumes. This will be discussed in a later section on differentiated service.
- *The possibility of excess entry*: It is complex to evaluate the net impact of entry under network diversity. The entrant's revenues can be represented as demand created or demand diversion, and also the extent of the welfare gains that result from providing network services that better satisfy the customers cannibalized from the currently existing network service providers.
- *The transaction costs of network diversity*: Introducing network diversity definitely increases transaction costs. For example, the network providers may have to upgrade their network to be compatible with the new standards. For the same reason, maintaining net neutrality will introduce extra transaction costs as well.
- *Long-run dynamic efficiency gains versus short-run static efficiency losses*: The entry of new competitors is not instantaneous, therefore the delay effect, which the market has to suffered short-run static efficiency losses is inevitable, even in the assumption that monopolistic competition provides the benefits of dynamic efficiency in the long run.

Another term that may be confused with net neutrality is fairness. Fairness is a prevalent issue in many aspects of life, and there is a significant literature on fairness in the area of data network. Jain in 1984 firstly proposed a quantitative measurement of fairness in communication networks [46], and such measurement was named as Jain's fairness index afterward. Later, Kelly [47] and Massoulié [48] further extended the fairness criterion. Zukerman *et al.* [49] proposed the α-fairness index. Researchers are mainly focused on the study of fairness between various transmission flows, especially the allocation of transmission bandwidth between different TCP (Transmission Control Protocol) flows [50–55]. Net neutrality is different from the criteria of fairness defined in data networks, as it restricts the ISPs' behavior, rather than specifically define resource allocation among users. Any fairness criterion is much more specific and well-defined than the net neutrality principle.

To simplify the explanation from a policy perspective, net neutrality can be explained using the following terminology:

- Vertical Integration
- Zero-price
- Price discrimination
- Zero-rating
- Non-discrimination

to be discussed in detail next:

2.2.1 Vertical Integration

The term "Vertical Integration" originates from economic theory, and describes an arrangement in which the supply chain is owned by the same company. It is considered that vertical integration exists between retailers (ISPs) and manufacturers (CPs) in the broadband industry.

Vertical integration in product line In the broadband service industry, companies that produce web-page and videos, provide e-commercial services and other applications are considered as a manufacturing stage. On the other side, DSL providers, Fiber-To-The-Home (FTTH) broadband service, mobile broadband service, and other last-mile technologies comprise the retail stage. One example of “vertical integration” in Internet access is a cable TV company that provides end-users Internet access together with their charged TV content, such as live-sport broadcasting. Vertical integration theory emphasizes that: “any vertical chain of production will only be efficient if every link is competitive.” To make a network environment fair and neutral, network diversity must be guaranteed. Yoo in [45, 56] examines the welfare implications of network diversity and concluded that “whether or not network diversity would promote economic welfare is an empirical question.”

To analyze the impact of incentive investment with/without net neutrality, Choi Pil and Kim in [57] proposed a model, where there exists a monopolistic service provider together with two content providers. They concluded that under net neutrality, vertical integration does not affect the service provider’s resource allocation in either the short term or long term. They further pointed out that social welfare does not decrease due to the prioritized service in a discriminatory network, where vertical integration may influence ISPs’ incentive of investment. A detailed economic analysis will be given in the later section.

Vertical integration in layered model It is worth to note that, as network service consists of “layers” which can be interpreted as (from the bottom to the top) [58]: (1) Physical/Instruction Layer; (2) Logical/Code Layer; (3) Application Layer and (4) Content Layer. Any product, provided by the service provider, intersecting these layers can be regarded as vertical integration. For example, Microsoft develops its own web browser “Edge”, and includes it in its products such as the operating system “Windows” and the online game console “Xbox One”. In this case, “Edge” is content and “Windows” is an application, and “Xbox” can be regarded as infrastructure.

2.2.2 Zero-pricing

During the development of the Internet in the 1990s, one-side pricing was the most common pricing scheme in the broadband market. However, in the mid-2000s, a few ISPs began to charge large content and application providers extra fees to reach their customers [59]. Such extra charge was named "termination fee".

In the literature of net neutrality, the zero-price rule was firstly proposed in [60], and discussed in [3, 61]. According to this rule, ISPs are not allowed to charge CPs for transmitting data to end-users. However, in certain circumstances, ISPs may have the incentive to weaken the competitive prospect of a CP, in order to favor a rival CP with which the ISP has a business relationship. To clarify, CPs, like end-users, must pay fees for their own connectivity to the Internet via ISPs (e.g. yearly or monthly according to the contract signed between the parties), but the zero-price rule forbids ISPs to charge CPs additional fees for transporting their data (e.g. per bit) to end-users.

The zero-price rule was proposed to deal with the following two strategies of any service provider of a platform [60]: exclusion and extraction. Exclusion is the action where the ISP favors one CP to induce its rival's exit from the market. For example, one ISP favors CP_A, from which it receives, say, higher payment, but discriminates against CP_B by throttling the bandwidth of CP_B's data transmission. Then, CP_B will eventually go out of business with its consumers ending their subscriptions. Extraction is the action that an ISP imposes charges to obtain a share of the CP profits. An instance [24] of extraction is when one ISP charges CPs for fast and stable access (in terms of premium access) of their content. The zero-price rule forbids both exclusions and extraction.

The key reason for the zero-price rule is welfare-enhancing: Consumers value additional CPs more highly than CPs value additional consumers. Without such a rule, ISP will stifle smaller CPs, favoring the stronger CPs that pay more, and as a result, the overall number of CPs and choices to customers and competition among CPs will be reduced. Weisman *et* *al.* [61] further pointed out that the zero-price rule excludes differential pricing, hence avoid prioritized service for CPs, and economic welfare will be reduced as a result.

2.2.3 Price-discrimination

Price discrimination is a well-known economic concept; such a practice exists when sellers are allowed to charge different prices to different buyers. Price discrimination is a selling strategy where a seller tends to charge each customer the highest price which he/she is willing to pay. In practice, price discrimination may change its form so that sellers divide customers into different groups with different prices according to certain properties such as customers' age.

Because elasticity of demand exists in the sub-market, price discrimination is adopted when more profits are made for sellers in sub-divided markets compared to an integrated market. A few manufacturers even intentionally damage a portion of their goods to achieve price discrimination [62]. Generally, there is an inversely proportional relationship between elasticity and price, i.e., customers in higher elastic sub-markets are charged lower price while customers in relative inelastic sub-markets are charged a higher price. In practice, companies divide the market into sub-segments with different elasticity. Such sub-segments can be insulated by nature of use, distance and time.

Basically, price-discrimination can be classified into three degrees:

- First-degree: The seller charges the maximum price as much as it can for each consuming product. In this case, prices vary among units, and the seller takes all possible consumer surplus for itself.

- Second-degree: The seller charge different prices for different quantity of goods, usually the greater the quantity, the less the price.

- Third-degree: The seller charges different prices for different groups of customers. For example, a railway company offers a half discount for the elderly and children.

In telecommunication and data service networks, price-discrimination can be interpreted as other forms such as free-charge or discount for certain websites or contents, a very prevalent situation in mobile data networks. This will be discussed in detail in the next subsection. In wireline networks, Weisman and Kulick [61] pointed out that price discrimination may have two forms in the relationship between ISPs and CPs: To enhance CPs' QoS such as prioritized access to their end-users (refer to the Netflix's fast lane [63, 64]), ISPs are allowed to charge uniform premiums. On the other hand, for the same level of service, ISPs would charge different prices to different CPs.

2.2.4 Zero-rating

Zero-rating is different from zero-price. It means that under certain cases, the consumers are able to access certain websites, services, or applications without being charged. It is also called "toll-free".

Zero-rating is usually regarded as a sub-topic of net neutrality and mainly arises in mobile data services [65–69], since the data volume in mobile data networks is more sensitive than wireline networks. An example is Wikipedia's website is free to access when using mobile devices in some countries. Zero-rating in wireline networks was reported in [70]: Netflix's CEO Reed Hastings criticized that Comcast did not count the data caps when its subscribers watch TV services through an Xbox.

Internet services such as Wikipedia, Google, and Facebook use "zero-rating" to promote their services in the emerging markets of most developing countries, where most mobile end-users cannot afford relatively expensive mobile data packages. However, such zero-rating is regarded as strictly against the net neutrality rule. It is worth noting that net neutrality proponents are split on zero-rating, which may greatly promote the Internet in developing countries [71].

Marsden [7] comparatively examined the implementation of net neutrality regulations on the national level, focusing especially on zero-rating. The case study part includes South America North America and Europe areas from 2003 to 2015. It is so far the most comprehensive summary work of zero-rating. In this work, the author separated and explained net neutrality into "negative" and "positive". Table 2.1 list the publication of regulation of "zero-rating" in different countries chronologically.

Country	**Legislation/regulation**	**Published**	**Date Enforced**
Norway	Guidelines	24/2/2009	Zero-rating declaration by NKOM of 2014
Costa Rica	Sala Constitucional De La Corte Suprema De Justicia	13/7/2010	2010 by Supreme Court precedent
Chile	Law 20.453'	18/8/2010	Decree 368, 15/12/2010
Netherlands	Telecoms Act 2012	7/6/2012	2014 and Guidelines 15/5/2015
Slovenia	Law on Electronic Communications 2012	20/12/2012	Zero-rating 2015
Finland	Information Society Code (917/2014)	17/9/2014	2014
India	Regulations (No.2 of 2016)	8/2/2016	August: 6 months after Gazette publication date
Brazil	Law No. 12.965	23/4/2014	Consulation 2015-16, no implementation

Table 2.1 Notable net neutrality laws or regulations (Source: [7])

Based on [7], a world-wide zero-rating case study is listed in the following:

- Norway: The Norwegian net neutrality guidelines were launched in 2009. Stakeholders held an annual meeting to examine the net neutrality status in Norway. It is not considered net neutral when ISPs send content via dedicated lines or use prioritized level services under a “best-effort” transmission mechanism. An electronic commerce law that took effect in 2013 stipulated that ISPs must follow the guidelines to ensure that no “self-regulation” happened. In 2014, zero-rating was defined as violating the net neutrality principle.

- Netherlands: Net neutrality in Netherland was triggered by service guarantees without discrimination. Customers complained that ISPs throttle the traffic of the messaging service WhatsApp. On March 6th, 2012, the Dutch Senate voted for net neutrality and finally became the first European nation to mandate a network neutrality law. As a consequence, the practice of zero-rating was banned in the 2015 Guidelines of the 2012 law. The investigation revealed that such prohibition only affects mobile ISPs.

- Slovenia: The net neutrality law in Slovenia is Article 203 of wider Electronic Communications Law 2012 (ZEKOM), which took effect at the start of 2013. This law regulated mainly against zero-rating. All major ISP in Slovenia were instructed to stop providing zero-rated service: Telekom Slovenije zero-rated for video channel HBO and UEFA Champions League football, and music streaming service Deezer. Si.mobile, the largest mobile ISP in Slovenia zero-rated for cloud storage service Hanger Mapa. AMIS and Tušmobil (Two ISPs in Slovenia) were banned to provide zero-rated mobile TV service.

- Chile: Chile legislated net neutrality on 10 August 2010. Four major ISPs in Chile were notified to stop zero-rating in 2014. However, it was misreported in media that all zero-rating service had been banned in Chile since 1 June 2014. Originally, it was intended to stop the zero-rating practice of social networks such as Facebook (Internet.org. [72]), since the regulator concluded that social network apps are prone to benefit when the zero-rated offer is allowed. On 22 September 2014, Wikipedia

Zero [73] negotiated an exemption from the ban based on the reason that Wikipedia is neither social nor commercial.

- Brazil: Zero-rating was a common practice among ISPs in Brazil before 2014. In 2013, the surveillance of President Roussef's personal communications led to net neutrality being legislated in Brazil. In the beginning, the law did not regulate the zero-rating service by ISPs. Although zero-rating was debated during the consultation of net neutrality law, zero-rating is not totally banned at the time of writing [74].

- India: Zero-rated service was offered by Internet.Org (owned by Facebook) and Airtel (The largest mobile ISP in India). In a consultation of net neutrality in the spring of 2015, over one million replies focused on zero-rating. The regulator, Telecommunication Regulatory Authority of India (TRAI), released its recommendations on 28th November 2017, to enforce net neutrality nationwide in India.

- Canada: Canada issued its net neutrality rules in 2009 and enforced them until recently. The zero-rating practice in Canada is not very common but not banned. To regulate the zero-rating is part of the net neutrality battle, which is carried out as a broadcasting ownership battle in Canada.

- USA: In the United States, Congress will vote to roll back the FCC's repeal of net neutrality law issued by the Obama administration at the time of writing. In the net neutrality law, FCC offered three "Bright Line" rules: No Blocking, No Throttling, and No Paid Prioritization. This law will finally prohibit zero-rating, though zero-rating is a common practice nowadays in the United States.

- European Union: The legislation of net neutrality in the European Union was complicated since there are 28 member countries in the union, some of which had already issued regulations for net neutrality such as Norway, Netherlands, and Slovenia. As pointed out by the European Commission's memo-15-5275, zero-rating is a commercial practice by certain ISP (especially mobile operators), and can also be called sponsored connectivity. Zero-rating is harmful to competition in the access market since it reduces the choice of end-users. This memo further argued that net neutrality regulation rules must ensure every consumer's choice of their Internet access provider, and provide non-discriminatory traffic management.

2.2.5 Non-discrimination

For a long time, telecommunication networks have been categorized as common carriers with non-discrimination requirements. With the help of telephone networks especially the Digital Subscriber Line (DSL) technology, the Internet has been greatly popularized during the 1990s to 2000s. In 2005, the FCC reclassified Internet services from "Telecommunication services" to "Information services", where the non-discrimination rule is not compulsory [75]. This reclassification enables ISPs to offer prioritized services, allowing for Netflix and Hulu to pay ISPs extra money to let their content have higher priority over other data traffic.

Compared to the concepts and terminology discussed above, the "non-discrimination" rule is used much frequently in the debate of net neutrality. It means that all data packets sent to the network which bypass the ISPs are treated equally and that no intermediate node can exercise control over the network as a whole. Schuett in [76] pointed out that emphasizing non-discrimination rule in broadband service will restrict ISPs to offer a single product line. And such non-discrimination restrictions will lead to two types of response on the part of the ISPs. The positive one is that ISPs will offer different network access qualities with different prices, allowing CPs to choose the quality based on their own requirements. The negative one is that ISPs prioritize or degrade traffic without transparency, which means that ISPs will not reveal how their customer CPs' traffic is treated.

The non-discrimination rule represents a historic and romantic view of the Internet but neglects that the QoS requirement is always an important issue of data networks [26, 77]. In fact, the Internet Protocol version 4 (IPv4) already contained a Type of Service (ToS) field, located in the header of the IP packet. This ToS prefix allows services with different QoS requirements to be supported. However, since there is no general agreement on how to handle data with different ToS entries, the ToS field is almost unused. Researchers then successively proposed and developed integrated service and differentiated service architectures, as well as many other approaches to achieve services with different QoS levels.

In the next chapter, a detailed technical part of implementing differentiated services will be given.

2.2.6 Summary of Policy Perspective

We have discussed certain concepts and principles, with related practical examples, for understanding the policy implementation of net neutrality. Net neutrality proponents insist that the Internet should treat every packet "equally" without any priority, yet researchers have already considered setting up a series of standards and frameworks to support differentiated services to meet different QoS requirements.

As a summary, we can draw a figure to show the relationship between these concepts and principles. Fig. 2.1 briefly shows the structure of the taxonomy of net neutrality; it can be seen that zero-rating is regarded as a subcategory of price-discrimination.

Zero-rating enables entering the Internet without cost in some circumstances, which is another form of killing competition. No matter how helpful it is, the zero-rating subsidized by a few big companies is anti-competitive. Although there might be some positive effect in raising the Internet's penetration rate in the developing country, zero-rating will harm the broadband market in the long run. From case studies, one can see those big content companies, who are prone to net neutrality rule, are willing to promote their service through

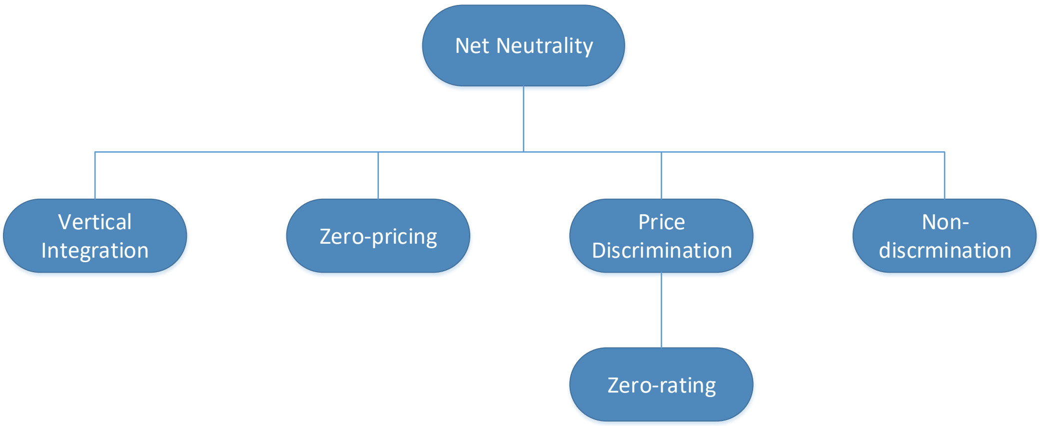


Fig. 2.1 A taxonomy of Net Neutrality form policy perspectives

the "zero-rating" policy in some developing countries, while such practice is regarded as being strictly anti-net neutrality rule.

Although legal (and even moral) aspects play an important role in making public policy, economic analysis and pricing models can give quantitative measurements to easily understand how the regulatory policies affect the broadband service markets, fundamentals of economic and pricing models of ISPs. This will be presented in the next section.

2.3 The Economic and Pricing models of Current Broadband Market

This section discusses net neutrality from an economic perspective. Economic models of net neutrality focus mainly on modeling Internet access networks where decisions are made on service differentiation and access prioritization. Such models help the industry understand the economic relationships between different stakeholders.

2.3.1 Visualized Models of Broadband Market

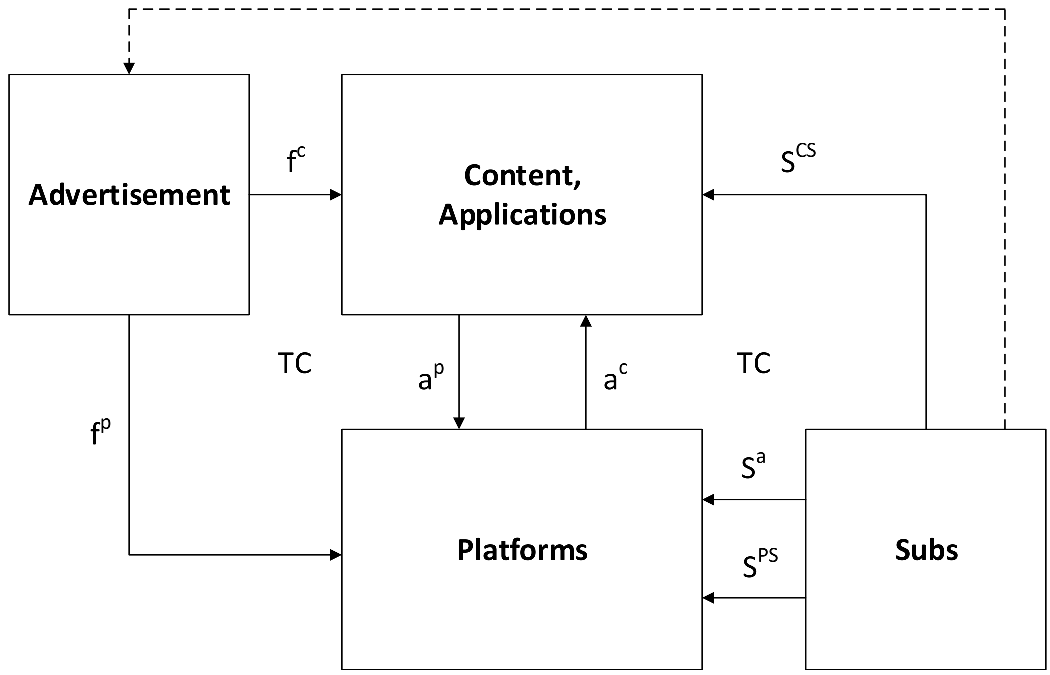
Given the massive size and complexity of the Internet access market, simplifications have been made to help better understand the economics of broadband access. For convenience, figures will be used to describe and visualize payments and monetary flow in the Internet access market.

A stylized pricing model of stakeholders in the broadband industry was proposed in [1], as shown in Fig. 2.2 (redrawn based on [1]). This model contains ISPs, CP/CAPs (Content Provider/Content and Application Provider), end-users, and also advertisers, who contribute the majority of the revenues of CP/CAPs and end-users. Not all the payments or revenues are applicable in the model. All the costs are assumed to be fixed, and incremental costs are neglected. In the figure, “Platforms” represents ISPs. “Subs” represents the end-users.

In this model, the parameter S^a stands for the access payment from subscribers to ISPs, and the parameter S^{PS} is service payment from subscribers to ISPs. This payment may be not applicable to cable broadband services, but can be used in DSL (Digital Subscriber Line) technology and mobile networks whose usage fee is based on the traffic volume. Parameter S^{CS} is the service payment from end-users to CAPs, parameters f^c and f^p are the advertising fees from advertisers to CAPs and ISPs, respectively. The parameters a^P and a^C are the mutual access prices of ISPs and CAPs, respectively. However, the access price a^C is often not applicable in practice.

Hermaline *et al.* [2] discussed and analyzed the broadband industry as a two-sided market. A two-sided market model is shown in Fig. 2.3.

In Figure 2.3 (i), “Platform” represents ISPs; it offers Internet access service to both the left-side “Application” and right-side “Household”, where “Application” represents CPs and “Household” represents end-users. ISPs get revenue $p(q)$ from CPs and h from end-users, respectively. End-users also pay the service fee t to CPs, and in the Fig. 2.3 (ii), CPs get the revenue a from “Advertisers”.



S^a … subscriber access payments

S^{PS} … subscriber service payments

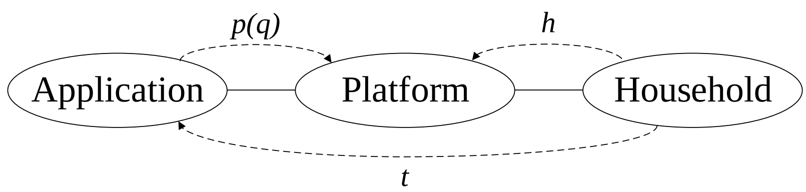
S^{CS} … subscriber service payments to C/A

f^c , f^p … payments by advertisers to C/A, P

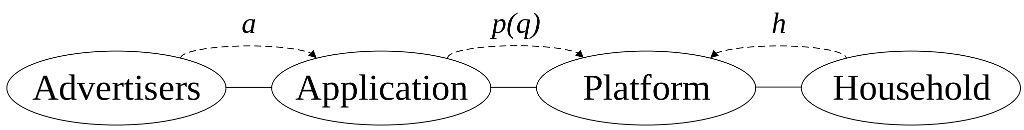
a^p, a^c … platform, content access payments

TC … transaction costs

Fig. 2.2 A stylized cost model of network [1]



(i) Fee for service model



(ii) Advertiser-supported-content model

Fig. 2.3 Two-sided model of broadband economic (Source: [2])

Musacchio *et al.* in [3] further assumed that ISPs are allowed to charge CPs which are not directly connected to them when sending data to their end-users. In Fig. 2.4, “A” stands for advertising agency (agencies), “C” denotes content providers, “T” denotes the ISP, and “U” represents the end-users. Dot lines with an arrow indicate that the payments are made under a two-sided pricing rule, which means that CPs need to pay the ISP when they reach that ISP’s subscribers. In this case, such pricing is regarded as “non-neutral”.

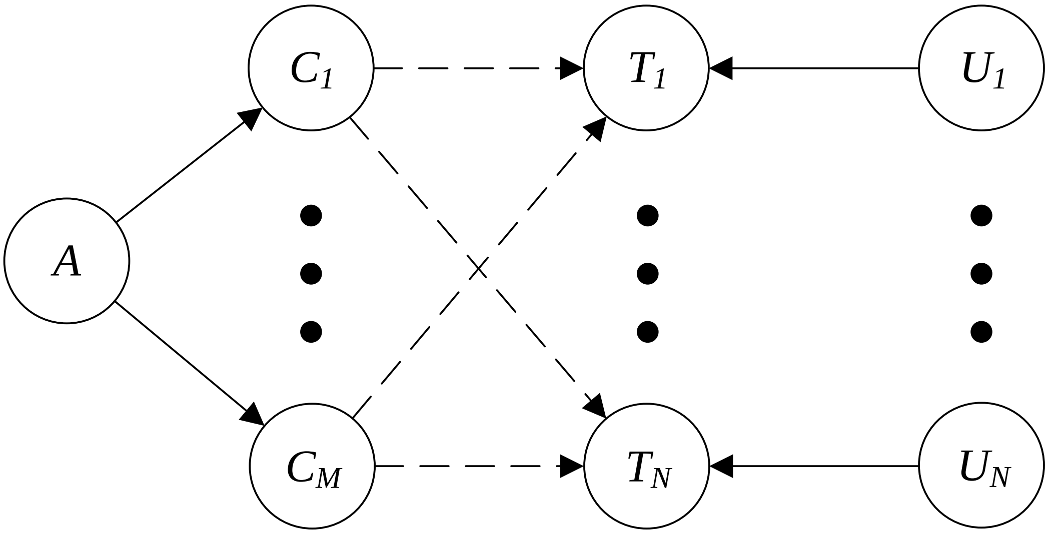


Fig. 2.4 The direction of payment of two-sided model (Source: [3])

2.3.2 The Pricing Model of Broadband Industry

The backbone of the Internet consists of tens of thousands of autonomous systems (ASs), where ISPs are attached, as shown in Chapter 1, Fig. 1.1. Given such complexity, the backbone of the Internet remains unregulated [78]. The pricing schemes between two ASs is usually based on bilateral settlements [79, 80]. Since the issue of network neutrality is concentrated in the last mile [81–84] of the Internet, the regulation inside the Internet backbone and related pricing schemes is beyond the scope of this work.

The research of the pricing model of the Internet originated from the 1990s when the Internet was evolving quickly. The authors in [85] first proposed to use “congestion pricing” to encourage more efficient usage of network resources. Gibbens *et al.* [86] proposed forming differentiated sub-networks to generate services class according to congestion levels, which are determined by the capacity of the sub-network and number of users therein. Other economic models of the Internet can be found in [87–94]. In the following, we will focus on highly cited pricing models adopted at the ISP level.

Paris Metro Pricing model

The Paris Metro Pricing (PMP) scheme [95] was inspired by the Paris Metro system, and therefore named after it. The basic idea of PMP is the following. Divide the metro into two classes, namely 1*st* class and 2*nd* class. The only difference between the two classes is that the 1*st* class costs twice as much as the 2*nd* class, which implies that the number of seats and comfortability of the two classes are the same. The expected result of this scheme is that the 1*st* class is quieter and less congested than the 2*nd* class. The PMP scheme has the self-regulating property, which can be explained like this: When the 1*st* class coach becomes congested, part of passengers feel it is not worth paying extra for 1*st* class, and so they move to 2*nd* class coach. Consequently, the congestion in the 1*st* class coach is mitigated, and the differential QoS between the two classes is therefore restored.

By applying the PMP scheme to broadband communications, the network will be divided into several logically separate channels, just like the 1*st* class and 2*nd* class coaches, each of which runs a "best effort" mechanism. If differentiated services are implemented, usage sensitive pricing will be needed. However, end-users do get used to flat-rate pricing, which refers to a pricing structure that charges a fixed fee for a service, regardless of usage. It is a norm nowadays for mobile ISPs to charge their subscribers a fixed price per month with limited mobile data usage.

The PMP scheme is intuitive and seems to be easily implemented. However, Gibbens *et al.* [8] concluded that the PMP scheme is not viable in a competitive environment with more than one ISP. The two competing ISPs' case is representative and we briefly explain the derivation of this conclusion of [8] as follows:

First, end-users' utility function $U(\theta, i)$ is defined. When a user chooses ISP i to connect to the network, benefit V of this network connection will be obtained, minus the product of the network's congestion parameter K^i and end-users' tolerance of congestion θ, and then subtract cost p^i paid by the user: $U(\theta, i) = V - \theta K^i - p^i$. The network congestion parameter K^i is defined as the total number of users Q^i divided by the total capacity of the network C^i: $K^i = \frac{Q^i}{C^i}$. Then the income of ISP i can be denoted as $\pi^i = p^i \times Q^i$. Each ISP aims to maximize its profit. To this end, ISP goes through the following steps: 1) it divides its network into subnetworks to provide different class services; 2) allocate capacity to each subnetwork; 3) set the price of each subnetwork.

At the second step, a Nash equilibrium may exists in a pair of price vectors: $(\mathbf{p}^{I*}(n^I, n^{II}),$ $\mathbf{p}^{II*}(n^I, n^{II}))$, such that for all n^I, n^{II}, and $\mathbf{p}^I(n^I, n^{II})$, where $\mathbf{p}$ is the price vector, n is the number of subnetworks of each ISP.

Intuitively, a network service provider with higher congestion charges lower price to attract end-users, i.e., $P^I \leq P^{II}$, when $Q^I/C^I \geq Q^{II}/C^{II}$. End-users who have a higher tolerance of congestion will join a cheaper but congested network, while other end-users will join the less congested network at a higher price. There is a critical value of end-users' tolerance θ^*, which is the equilibrium point of the price and congestion level. For each end-user, if $\theta \leq \theta^*$, $U(\theta, II) \geq U(\theta, I)$, users with lower θ will join network II. In the case that the two ISPs have capacities C^I and C^{II}, and they set their prices $P^I \geq P^{II}$, with critical value θ^*, the proportion of end-users who join ISP I is θ^*, and $1 - \theta^*$ for ISP II. Specifically, two competing ISPs with the same network capacity are analyzed and yield the following four propositions [8]:

Proposition 1 : Two ISPs have equal network capacity, and neither of them divides their network into subnetworks. A unique Nash equilibrium exists, where $P^I = P^{II} = 0.5/C$

and their profits are positive, $\pi^I = \pi^{II} = 0.25/C$. The level of equilibrium profits is a consequence of congestion.

Proposition 2 : Two ISPs have equal network capacity, one ISP divides its network into subnetworks, and the other did not. There is a unique Nash equilibrium in pure strategies: ISP II set its price $P^{II} = 0.4766/C$ where lies between the two prices set by ISP I: $P_1^I = 0.55784/C$ and $P_2^I = 0.45/C$. The profits of two ISPs are $\pi^I = 0.2455/C$ and $\pi^{II} = 0.2427/C$.

Proposition 3 : Two ISPs have equal network capacity, both ISPs divide their networks into subnetworks, and no Nash equilibrium exists.

Proposition 4 : Different from the above three propositions, proposition 4 conducts the subgame into stages, and the stage is symmetric for each network: The first stage does not do the subdivision. In the second stage, check the status of two ISPs, if none of them subdivides their network, set the pricing to 0.5/C, if ISP *I* did not subdivide its network, but ISP *II* subdivided its capacity into two subnetworks, ISP *I* charge 0.4766/C, if ISP *I* subdivide its capacity but ISP *II* did not, ISP *I* charges 0.5784/C and 0.45/C. If both ISPs formed two subnetworks, charge any two price (nonnegative) p_1^I, p_1^I for ISP *I*, and p_2^{II}, p_2^{II} for ISP *II*.

For clarity, Table 2.2 summarize the parameters' setting and results of each proposition.

Table 2.2 Four scenarios of duopoly case in [8]

Proposition	Network I					Network II				
I	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	N	C	0.5/C	0.25/C	Y	N	C	0.5/C	0.25/C	Y
II	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	Y		0.5784/C, 0.45/C	0.2455/C	Y	N		0.4766/C	0.2427/C	Y
III	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	Y	N/A	N/A	N/A	N	Y	N/A	N/A	N/A	N
IV	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	N	C	0.5/C	0.25/C	Y	N	C	0.5/C	0.25/C	Y
	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	Y		0.5784/C, 0.45/C	0.2455/C	Y	N		0.4766/C	0.2427/C	Y
	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?	Divide?	Capacity	Price	Total Profit	Nash Equilibrium?
	Y	N/A	p_1^I, p_2^I	N/A	N	Y	N/A	p_1^{II}, p_2^{II}	N/A	N

In these four propositions, numerical results show that when ISPs are free to choose the capacities of the subnetworks and the corresponding prices, competition is not sustainable when both ISPs want to maximize their profit, and equilibrium does not exist.

Chau *et al.* in [96] followed the analytical approach in [8, 86] where utility and congestion function are used as metrics. They concluded that for the capacity sharing service, PMP increases the profit of ISPs and increases the social welfare; for the latency based service, PMP decreases the social welfare but not desirable to ISPs since it reduces ISPs' profit. It should be noted that all the conclusion is based on achieving equilibrium, which means that end-users have the right to choose/opt out the service class in the process. However, in practice, end-users usually sign a service contract with their service provider; such contract specifies service terms (eg. access rate or data cap per month), and it is not feasible to make end-users shift service class in a short time (such as during the congestion period). Therefore, the PMP scheme may run well under a flexible service situation such as in a metro system, but its complexity increases when applying it to contracted service.

One- and two-sided Models

As mentioned in section 2.1.3, opponents of net neutrality feel unfair and are dissatisfied that two-sided pricing generates unbalanced revenues in the broadband market: ISPs receive only monthly fixed payment while CPs receive far more payment through advertisements and on-demand service (such as VoD). In the one-side pricing market, CPs and end-users only pay for their direct access. Specifically, the money that CPs paid to ISPs can be normalized as net of ISP connection cost, which means ISPs only charge end-users. Therefore, one-sided pricing is usually regarded as a "neutral" pricing scheme.

On the contrary, the two-sided market scheme allows ISPs to get revenue from both end-users and CPs: One ISP can charge all CPs, even certain CPs not directly connected to it. See Figure 2.5 for example. Local ISP T_3 can charge end-user u and CP C_1, even though

C_1 is not directly connected to T_3 for Internet access. Thus, two-sided pricing is regarded as “non-neutral”. The definition of the two-sided market was mentioned in [61, 97]. “A two-sided market can be thought of as a meeting place that brings together two distinct user groups, each of which benefits from the presence of the other.”

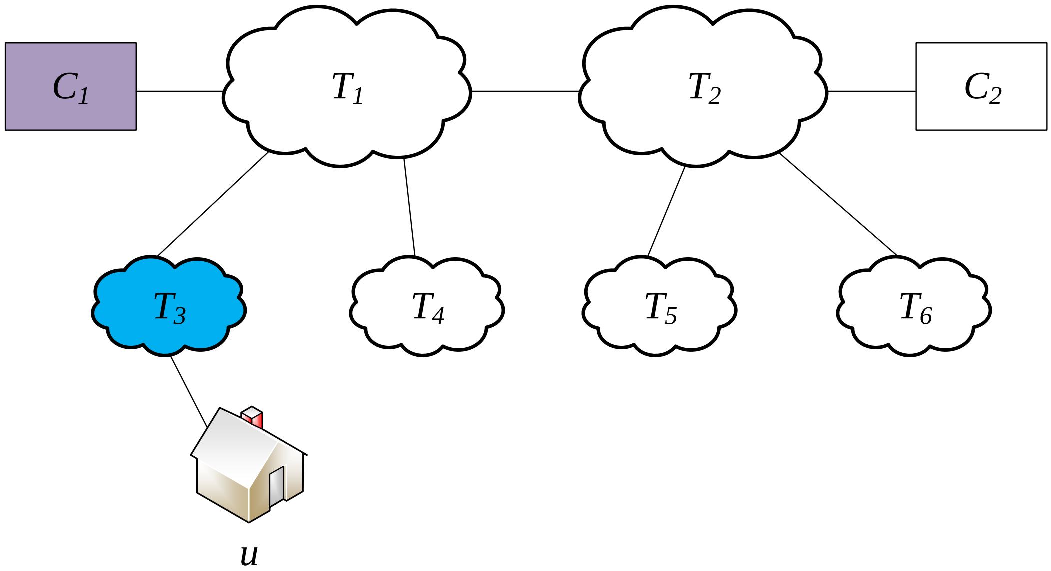


Fig. 2.5 Illustration of two-sided market (Redrawn based on [3])

The two-sided market model of the broadband industry is discussed and analyzed in [2–4, 57, 61, 98, 99]. The ISPs are regarded as the platform, and the CPs and end-users are modeled as two participants, respectively.

In the following, we discuss two issues associated with the two-sided model, namely single product restrictions and the incentive of ISP.

Single product restriction in broadband markets The net neutrality rule also guarantees traffic’s fairness, which restricts differentiated service (or paid prioritized) to be employed by each ISP. The effect of no differentiated service can be regarded as the “single-product restriction” in economics. Hermalin and Katz [2] discussed the single-product restriction in the broadband market, including monopoly ISP scenario and duopoly ISPs’scenario. In the monopoly ISP scenario, three consequences can be predicted: 1) end-users who consume the low-quality service variant before are excluded by the market; 2) the mid-class end-users consume a more efficient and superior quality; 3) the top end-users suffer a less efficient and inferior quality. Specifically, the end-users at the bottom, who have the lowest ability to pay and are typically intended to aid, suffer most by the single-product restriction. From the three consequences, we see that the first and third reduce welfare while the second one raises welfare. In a duopoly ISPs' scenario, without a single product restriction, two ISPs may offer full product lines engaged with head-to-head competition. However, introducing the restriction leads to identical products offered by two ISPs, which reduces the variety hence decreases the total social welfare. Overall, in both monopoly and duopoly cases, the single product restriction reduces the social welfare, therefore the consequence is contrary to the original intention.

The incentive of ISP in two-sided markets In communication networks, much research focuses on the "technology layer", which investigates the performance of networks including throughput and utilization. Walrand [4] pointed out that the "technology layer" should be combined with the "economic layer", where ISPs' investments affect network performance and network performance further affects how end-users utilize the network. Musacchio *et al.* [3] raised and answered the question that whether a "neutral" network should allow ISP to charge CPs when CPs transmit their content to end-users through ISPs' link (i.e., violate the "zero-price" rule)? As this issue is highly related to the investment incentives of both CPs and ISPs. Fig. 2.6 shows the investment flow in a two-sided market.

In order to achieve convenience and consistency, the analysis in [3] adopts the number of clicks, called "click rate" as the metric to characterize the usage of an end-user U_n. Click-rate is a natural measurement for an advertiser to pay for the CPs, but it is not an appropriate metric for counting ISPs' revenues from end-users as ISPs do not charge users based on their click rate. Annual or monthly payment is more commonly used by ISP to charge their subscribers (and is also more accepted by subscribers to pay). For the convenience of

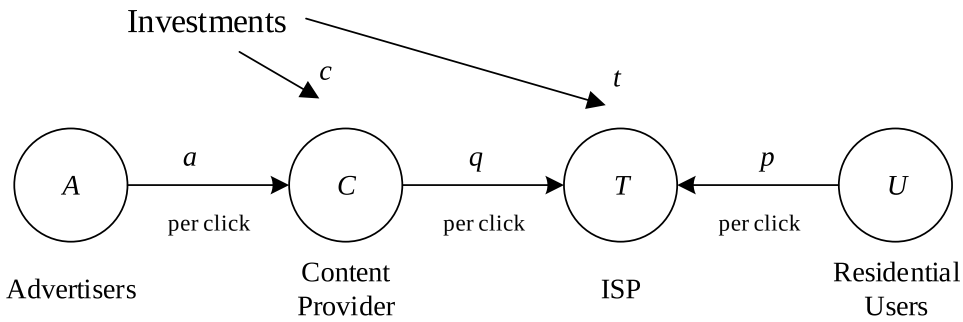


Fig. 2.6 The investment goes to CPs (C) and ISPs (T) (Redrawn based on [4])

calculations, only one "click-rate" metric is used. Following the notation and analysis of [3], the click rate of end-users depends on the investment of both ISPs and CPs, and can be expressed by:

$$B_n = \left\{ \frac{1}{N^{1-w}} (c_1^v + \cdots + c_M^v)[(1-\rho)t_n^w \times \frac{\rho}{N}(t_1^w + \cdots + t_N^w)] \right\} e^{-p_n/\theta}$$

where parameter N is the number of ISPs, and M is number of CPs. The term $c_1^v + \cdots + c_M^v$ is the investment of CPs, and the term $t_1^w + \cdots + t_N^w$ is the investment of ISPs. Parameter ρ represents the "spill-over effect"; when $\rho = 1$, the end-users U_n value each ISP's investment equally, while $\rho = 0$, the end-users U_n only value the investment of their local ISP T_n.

The investment of ISPs, such as increasing network capacity by deploying more optical cable, is regarded as a long term investment, while the investment of CPs, such as making content and improving the searching algorithm, is regarded as a short term investment. The analysis assumes ISPs choose their price (make strategy) first and that the content providers follow, which means ISPs' investment can be observed by all CPs, and CPs make their connection based on the ISPs' investment afterward, it is meaningless to reverse the play order of ISPs and CPs.

Based on these assumptions and models, one- and two-sided pricing are compared. The authors concluded that the best investment incentive is that revenue can be shared between

ISPs and CPs. Specifically, if CPs get more income than ISPs, it is desirable for CPs to share part of the revenue with ISPs and vice versa.

Choi and Kim [57] analyzed the effects of net neutrality regulations on investment incentives both for ISPs and CPs, and their implications on social welfare. Two types of networks are compared: one where the network allows prioritized operating, so-called "multi-tiered network", and the other one is a nondiscriminatory network under strict net neutrality regulation. The comparison showed that investment incentives depend on the tradeoff between the fixed monthly payment by end-users and revenues from CPs who pay for the premium delivery with high priority. The author pointed out that it is difficult to draw clear conclusions because the investment incentive interacts with net neutrality regulation subtly.

Other Pricing Models

PMP, one- and two-sided models are the pricing models that appear most frequently in research papers on broadband network economy. Besides these two models, other models focused on other perspectives have also been proposed.

In [100], the author modeled a bandwidth sharing mechanism. A pricing scheme which charge end-users based on their bandwidth usage was also proposed. The traffic is divided into classes to form a service differentiation. The optimal bandwidth for each user and the related minimum QoS in terms of call blocking probability required are then formulated. An optimal value of reserved minimum bandwidth can be achieved that maximizes the ISP's revenues for both single- and multi-classed scenarios.

A more recent work [101] revealed that the elasticity of system throughput plays a key role in determining the optimal pricing. Specifically, the capacity of ISPs and the delay sensitivity of users caused by network congestion are determined by the type of traffic (such as video and text). As video traffic has grown rapidly in recent years, ISPs have the incentive to adopt two-sided pricing. From the perspective of regulation, regulators might put more effort to regulate the price when ISPs have higher market power in a two-sided market.

2.3.3 Summary

Visualized models reveal the monetary flow between different parties such as ISPs, CPs, and subscribers (end-users). Network analysis from an economic perspective derives two pricing models, namely the PMP model and the two-sided model. PMP is a simple model that describes a tiered pricing scheme. As discussed in the 2.3.2, under certain assumptions, the PMP model may partially reveal the decision results of two competing ISPs, but it cannot fully explain the impact of network neutrality regulatory policies. Analysis of the two-sided pricing model is able to derive the investment incentives of ISPs and CPs.

2.4 Performance Analysis and Evaluation

From an engineering perspective, network neutrality issues can be understood as allocating transmission capacity, while considering both equity and efficiency. Queueing theory and game theory are two commonly used analytical tools. Queueing theory deals with the congestion problem, while the game theory is used to study the benefit distribution among different stakeholders.

The problem of congestion is intrinsic in any network: it happens when the transmission capacity of a particular route (or channel) is lower than the transportation demand. Such congestion may be incurred by the users' self-optimized strategy of choosing routes (e.g. shortest path routing). It may achieve equilibrium in the local network but lead to an inefficient system state. Queueing models are often adopted to analyze congestion dynamics.

2.4.1 Basic queueing model and its application in differentiated services

In [57] and [102], the M/M/1 queueing model is adopted to analyze the congestion problem. The reason for using M/M/1 model lies in two reasons: First, the cause of net neutrality debates comes from the contradiction between scarce transmission resources and the increasing demand generated by bandwidth-intensive service over the Internet. Second, this setting is a well-accepted approximation when the number of customers is sufficiently large and the service duration of each customer is independent.

In a network which treats all data packet without any priority (a.k.a "neutral network"), the expected waiting time of each customer can be expressed as:

$$w = \frac{1}{\mu - \lambda}$$

where λ is the average arrival rate of the network, μ is the processing speed of the network. This expression assumes that $\mu > \lambda$, otherwise the delay is negative, which is practically impossible.

When considering service tiering, let the arrival rate of the first priority be λ_1, i.e., the highest priority, which preempts all the other priorities with the lower level. The expected delay of first priority can be expressed as:

$$w_1 = \frac{1}{\mu - \lambda_1},$$

Combining the above equations, if there are only two classes of service, customers other than first priority experiences the expected waiting time is:

$$w_2 = \frac{\mu}{\mu - \lambda} w_1 = \frac{\mu}{\mu - \lambda} \frac{1}{\mu - \lambda_1}.$$

One can find that when adopting two classes of service tiering, high priority will get lower average delay while the lower priority will suffer higher average delay: $w_2 > w > w_1$ for $\mu > \lambda$. Further one can find the relationship between w_2 and w_1 to be: $w_2/w_1 = \mu/(\mu - \lambda) > 1$.

2.4.2 Congestion equilibrium in the last mile of the network

The network tries to meet the traffic needs of its users. Conversely, the degree of congestion on the network will also affect changes in user demand. In [103], Ma and Misra studied two types of congestion: CP-level congestion and user-level congestion and further derived the equilibrium states of these two types of congestion, respectively. To study these two types of congestion, three parties are modeled: 1) ISP located at the last mile, 2) CPs, and 3) end-users.

First, a monopoly ISP scenario is considered, and the model is as follows: Parameter μ is the capacity of the last-mile ISP, which connects a set of CPs with a number of N, parameter λ_i is the CP_i's aggregated throughput, hence the total throughput of CPs is: $\lambda_N = \sum_{i\in N}\lambda_i$. Parameter $\hat{\theta}_i$ is defined as unconstrained throughput for CP_i. Without contention, CP_i's throughput can be meet, i.e., $\lambda_i = \hat{\lambda}_i$. However, when $\mu < \sum_{i\in N}\hat{\lambda}_i$, the actual throughput of CP_i is less than its traffic demand $\hat{\lambda}_i$: $\lambda_i < \hat{\lambda}_i$. A nonnegative parameter $\phi \in \mathbb{R}_+$ represents the congestion level, then the throughput rate of each CP can be expressed as a function of number of end-users and congestion level:

$$\lambda_i(M, \phi) = \alpha_i M \rho_i(\phi_i)$$

where $\rho_i(\phi)$ denotes per users achievable rate, or demand function from end-users for each CP i, $\alpha_i \in [0, 1)$ is the percentage of end-users access CP i.

Based on this setting, the congestion level Φ of an ISP then can be represented as a continuous function of the throughput Λ of all the CPs and the capacity μ, $\Phi = \Phi(\Lambda, \mu)$.

Accordingly, the congestion level φ of a end-user can be represented by $\varphi = \varphi(M,\mu,N)$. The user-level congestion equilibrium is then defined through the fixed point equation:

$$\Phi(\Lambda(M,\phi),\mu) = \varphi$$

2.4.3 Net Neutrality Impact Analysis

ISPs, CPs, and end-users naturally form the three parties of a game. As a result game theory has been extensively adopted in the analysis of net neutrality issues, especially in the benefit/welfare analysis.

Richard T. B. Ma and his research team have undertaken extensive research work [79, 80, 103–112] on the net neutrality problem by using ideas from game theory and other related tools.

To give a chronological structure and a clear view, we summarize the related literature in the following tables 2.3, 2.4, 2.5:

2.4.4 Summary

The M/M/1 queueing model derives the QoS boundaries of different service classes, and can further derive the equilibrium state between user-level congestion and CP-level congestion. The game-theoretic analysis allows ISPs, CPs, and users to get the information benefit gained under different network settings (whether or not to adopt network neutrality policy). As emphasised by Misra in [13], competition in the last mile is a more important issue than the net neutrality regulation rules. From tables 2.3, 2.4 and 2.5, one can conclude that differentiated services is welfare enhancing when fully competition is guaranteed between ISPs.

People from the engineering field seem to be more inclined to implement differentiated services on the Internet. Davies et al.[114] suggested that network operators develop networks capable of providing different levels of service quality for various traffic classes. Crowcroft in [115] claimed that there was no “network neutrality” in the past, and that it is not necessary to engineer for it.

2.5 Summary

This chapter has comprehensively surveyed the issue of “Net Neutrality”. The evolution of the debate and key events reflect the complexity of the issue.

From the regulator’s perspective, some rules need to be established to ensure fair competition and prevent public interest from being compromised. From the perspective of operators, the principle of network neutrality should not hinder the normal profit of network operators. From the perspective of network users, the principle of network neutrality protects their right to use the network properly. As users’ demand for network bandwidth is constantly growing, how to ensure that network operators’ investment motivation for new network bandwidth is related to the greater well-being of the entire society.

The focus of the debate is whether network operators are allowed to provide differentiated services. Although many of the research papers are based on ideal assumptions, most of them (especially economic and engineering) have shown that under full competition, differentiated services will improve users’ choice and enhance the welfare of society as a whole.

In summary, the penetration rate of networks is different in each country, and the distribution of network resources is uneven. Therefore, network neutrality is a problem that varies from place to place, and case by case. How to use the network to ensure both fairness and efficiency is still an issue worth exploring. Are there universal network neutrality regulations? Instead, we envision more flexible regulation rules tailored to local conditions.

Researcher(s), Year(s) and Literature(s)	Conclusions related to Net Neutrality Regulation	Remarks
Ma *et al.* 2007, 2010 [79, 104]	N/A	The authors use "Shapley value" to analyze the profit sharing between ISPs, this work is the preliminary of their subsequent research on net neutrality.
Ma *et al.* 2011 [80]	Shapley solution can be used as a pricing structure for differentiated service therefore to enrich the Internet service	The authors extend their previous work [79, 104], further classified the ISPs into three categories: Transit, Content, and Eyeball.
Ma & Misra 2012 [103]	Under certain parameters such as the number of end-users M, system (ISPs') capacity μ, and premium class ratio κ of ISP's total capacity, the number of CP who choose premium class P increases against the M and κ, but decrease against μ	The authors assumed that ISP divides its capacity into two classes: Ordinary class O and Premium class P, CPs can choose service class based on their own decisions. By using queueing theory, the congestion equilibrium in both end-users level and CP level in a monopolistic ISP scenario can be derived, which enables to analyze the feasibility of a non-neutral single ISP network.
Ma & Misra 2013 *et al.* [105]	In an oligopolistic ISPs scenario, fully competition between ISPs ensures that the result of ISPs' non-neutral strategies is inline with end-users' utility. A public option ISP could be deployed as a backup choice in case that consumers' utility is hurt by existing commercial ISPs' strategies.	As differentiated services in a single ISP scenario had been analyzed in [103], the author proposed an idea that using a public option ISP to resolve the dispute on neutrality in an oligopolistic ISP's scenario. A public ISP is defined as an ISP does not divide its capacity into different service classes or charge any CPs.
Tang & Ma 2014, 2019 [106, 113]	In a monopolistic ISP scenario, the profit-optimal strategy of the ISP hurts the utilities of consumers. To solve this issue, a more relaxed regulation framework was proposed. Under this framework, the worst performance of the ordinary class is bounded, and the CPs are influenced regardless of the ISP's capacity. The ISP still keeps the incentive to enlarge its capacity, and end-users' utilities are positively correlated with CPs' profit.	The authors considered a monopolistic ISP scenario, where the ISP adopts a PMP-type pricing mechanism.

Table 2.3 Summary of net neutrality literatures using game theory

Researcher(s), Year(s) and Literature(s)	**Conclusions related to Net Neutrality Regulation**	**Remarks**
Wang *et al.* 2014 [107]	Paid prioritization increases the incentive of an ISP to expand its capacity, although such incentive depends on the cost, system scale, and the growth of users' demand. A monopolistic ISP's profit-optimal pricing strategy is inline with social welfare and it is also good for differentiated CPs who have a higher value, the drawback is such differentiation sacrifices CPs who have a lower value. From the policy marker perspective, the price of prioritized service should be restricted to a limit to balance the fairness among different stakeholders and social welfare.	The authors focused on comparing social welfare with or without net neutrality regulations under a monopolistic ISP scenario.
Ma 2016 [109]	Although it is prohibited by the net neutrality principle, subsidization from high-value CPs to their end-users increases the revenue of ISPs and social welfare. ISPs' incentive of expanding their capacity is also increased. The drawback of such subsidization might be that ISPs will raise prices and increase the network congestion, therefore low-value CPs will suffer lower throughput.	As discussed in the policy part, prioritized service and two-sided market are not allowed under net neutrality. The author studied the issue that CPs with high value subsidize their end-users and its related impact on the broadband ecosystem.

Table 2.4 Summary of net neutrality literatures using game theory (continued)

Researcher(s), Year(s) and Literature(s)	**Conclusions related to Net Neutrality Regulation**	**Remarks**
Ma 2016 [108]	When capacity is scarce and users are not sensitive to network congestion, a monopoly provider may increase its profit through service differentiation (Paris Metro Pricing). When users value the service higher, the optimal pricing of providers could also be higher, in both monopolistic and duopolistic market. From the regulator's view, the author suggested that: 1) limit the monopoly provider's price to guarantee social welfare. 2) allowing service differentiation increases both social welfare and the provider's profit. 3) Competition reduces price, result in higher social welfare can be achieved.	Applying usage-pricing in a competitive market was studied in this paper. The author used a two-dimensional model to capture both the users' value v on per-unit service usage and demand sensitivity w to network congestion.
Ma *et al.* 2017 [111]	Compare to a neutral network, prioritized service (or service differentiation) increases social welfare. ISPs have a higher incentive to expand their capacities under no net neutrality regulations. The authors concluded that prioritized service in the network could be viable for the future.	This paper focused on the key concern about net neutrality: whether paid prioritization should be allowed in a monopolistic market.
Ma 2017 [110]	The users' loyalty and its price determine whether a CP use premium peering. Low-value CPs' peering action gives pressure on high-value CPs, but not vice versa.	The premium peering means CPs pay ISP for a prioritized or better connection rather than ordinary best-effort connection. Such practice is against the "non-discrimination rule" in the net neutrality principle. In fact, Netflix and Hulu used premium peering to get fast connection to reach their end-users in 2014, which had been mentioned in 2.1.
Zou *et al.* 2018 [112]	The authors analytically derived the ISP's optimal service differentiation strategy for its maximum profit. ISP's profit-seeking nature definitely hurt the welfare of low end-users, so competition might be introduced to give an alternative choice for low end-users.	Put aside the net neutrality principle, the authors focus on the issue that ISP's profit maximization in network capacity constraints.

Table 2.5 Summary of net neutrality literatures using game theory (continued)

Chapter 3

Differentiated Traffic in Data Networks

A number of key events have been reviewed, and the literature on net neutrality has been comprehensively studied in the previous chapter. During the evolution of the Internet, a series of mechanisms have been designed to offer differentiated services to meet different QoS. This chapter discusses the technical details about the mechanism and implementation of differentiated services in data networks.

3.1 The Motivation and Emergence of Differentiated Traffic in Data Internet

In many business areas, the concept of differentiated services is widely accepted, and is common practice. Typical examples are commercial airline tickets differentiated into classes by prices and differentiated mail/delivery service: the more customers pay, the faster the delivery.

The Internet, however, was not designed for offering differentiated services from its very beginning. The Internet was developed by the Advanced Research Projects Agency Network (ARPANET), and originated from the idea that sharing scarce computation resources among geographically distributed universities in the US from the 1960s to the 1970s. The basic principle of ARPANET is to ensure simplistic end-to-end communication; differentiated services was not being considered. Due to the limitation of technologies, text messaging was the main media being transmitted at the beginning. Ever-evolving telecommunication technologies such as increased transmission capacity have enabled the transmission of voice and video, and have brought out many protocols for different services. The following protocols, nowadays have become the standards: Simple Message Transfer Protocol (SMTP) is used for email transmission, Hypertext Transfer Protocol (HTTP) for web browsing, User Datagram Protocol (UDP) for connectionless data transmission such as Voice over IP (VoIP), and the most used: TCP for reliable end-to-end data transmission.

3.1.1 QoS in data networks

The term "Quality of Service" measures the performance of a network in numerical terms according to prespecified metrics, and indicates the statistical performance guarantee of a network [116–119].

In the traditional telephone system, a call is connection-oriented. The calling process requires two ends of the call to establish an end-to-end connection. Once the connection is successfully established, the resource of such a connection cannot be used by other entities or devices until the call is finished. Hence, maintaining the QoS of a telephone call is relatively easy: Ensure that the resource being used is not occupied by other call requests. However, in a data transmission network (e.g. Internet), the data is divided into data packets, each packet is assigned an IP address which indicates its source and destination. The TCP/IP suite is the most widely operated protocol on the Internet at present. The QoS of Data networks aims to quantify the process a packet can expect when it is being transmitted in a network. TCP/IP networks deliver packets in the "best effort" manner. In other words, it delivers packets as much as possible. TCP adopts an algorithm named AIMD (Additive Increase and Multiple

Decrease) [120, 121] to explore the available bandwidth through the link. However, this does not guarantee the QoS of each packet, in other words, the “Best Effort” mechanism does not allow resource reservation. Before QoS mechanisms were deployed, showing and indicating the guaranteed QoS was not a common practice for an ISP. Instead, ISPs announce service level agreement (SLA) instead of QoS, Reference [122] is an example of SLA published by Verizon in the U.S. SLA usually contains the terms including “Site Availability Scope” and “Time To Repair”, which means the websites that can be accessed when using the service provided by an ISP and the average time required to repair a failed component or device, respectively. Due to the lack of QoS at the packet granularity in SLA, performance metrics such as average delay in accessing a certain website are not available.

3.2 Approaches to implementing differentiated services

In this section, we discuss various approaches that implement differentiated services in data networks.

In today’s data networks, data needs to be packaged layer by layer, then passed through the transmission process, and finally presented to users. The Open Systems Interconnection (OSI) reference model was released by the International Organization for Standardization (ISO) in 1984, which explicitly divided the network into seven layers. The layers are named from the bottom-up: physical layer, data link layer, network layer, transport layer, session layer, presentation layer, and application layer, respectively. Fig. 3.1 shows the structure of this OSI 7 layers model and commonly used protocols therein.

The OSI 7 layer model is a concept that standardizes the function of two or more communication entities without regard to their underlying internal structure and technology, and aims to partition the communication system with the help of standard communication protocols. The use of a hierarchical structure has the following advantages: the hierarchy and components under the standardized definition enable the network to run on platforms

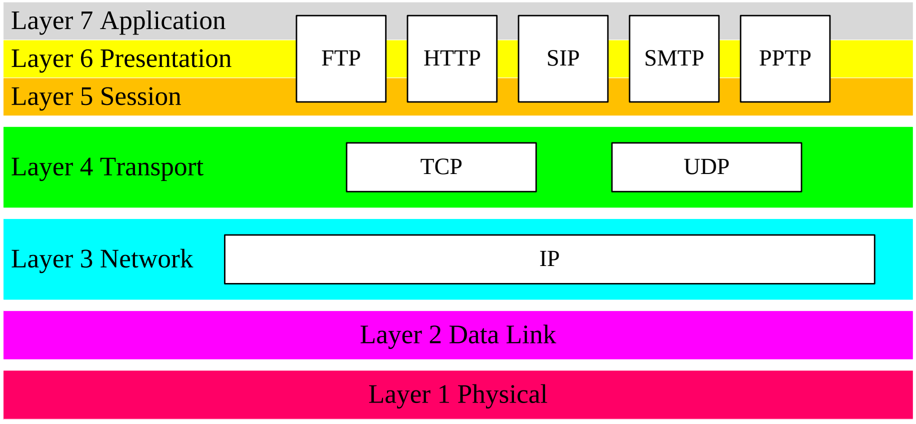


Fig. 3.1 The illustration of OSI 7 Layers model and commonly used protocols

built by different network providers. Because each layer is independent, the change of each layer does not affect the operation of other layers, the complexity of maintaining the network is simplified, and technological advances at each layer can accelerate the evolution of the network.

Under the OSI 7 layers model, the transmission of data needs to be layered and packaged. Theoretically, differentiated services can be implemented at each layer. Since the technical detail of each layer had been well documented, it is not worth repeating them here again.

Inspired by previous work [123–125], which study traffic optimization in the multi-layered networks, we will briefly discuss the principles and logic of implementing differentiated services in different layers from the perspective of a layered model.

3.2.1 Physical Layer

The physical layer is located at the bottom of the OSI 7 layers model and is responsible for the communication of raw data stream over a physical medium. The information in the physical layer can be categorized into three types: electrical signal, optical and electromagnetic waves, which correspond to co-axial cable, optical fiber, and wireless transmission medium, respectively.

Differentiated services in the physical layer can be realized by setting dedicated lines (a.k.a. "leased lines"). Such dedicated lines only serve certain exclusive customers. Take the operation in the public switched telephone network (PSTN) for example. A telephone line can be invoked by the switching station to establish a call, and released by the switching station when the call is ended. Hence, a new coming call could be blocked in the situation that the switching capacity is full, so that the QoS is degraded when the traffic is very high during peak hours. The leased line, in contrast, only connect two telecommunication terminals, and is not assigned by public switching station, and is never used by other users. Therefore, telecommunication by leased lines is more reliable and its QoS is relatively higher than that provided by public switching mechanisms.

For fixed-line networks, leased lines are in the form of coaxial, fiber optics cables. In wireless networks, leased lines can be implemented by beamforming techniques [126].

Using leased lines in the physical layer is an intuitive and reliable way to achieve differentiated services in the network. However, it is a relatively costly way that promotes low utilization of network resources. In practice, leased lines are far more expensive than public switched networks. Therefore, in order to reduce cost and improve the utilization of network resources, the implementation of differentiated services in the network is mainly concentrated in high-level protocols.

3.2.2 Data Link Layer

The data link layer contains two sublayers: the logic link control (LLC) and the media access control (MAC). The logic control layer takes charge of flow control and guarantees an error-free for accurate data transmission. The MAC layer manages data access into network nodes and transmits data between the network nodes. Differentiated services was implemented in the MAC sublayer.

Two Examples in Data Link Layer

Two examples including both optical networks and wireless networks are given next:

Optical local area networks: With relatively low cost and high capacity, optical communication is now widely used in both core networks and local area networks. Ethernet passive optical networks (EPON) are an ideal solution to local area networks. Kramer *et al.* in [5] investigated differentiated services in EPONs.

In their EPON model, an optical line terminal (OLT) connects a number of optical network users (ONU). As can be seen in Fig. 3.2, there are two types of priorities in the system, namely Intra-ONU scheduling and Inter-ONU scheduling, thus two levels of priority scheduling are implemented. The Multi-Point Control Protocol (MPCP) in IEEE 802.3ah defines and coordinates data traffic between ONUs (inter-ONUs), and IEEE 802.1D takes charge of the data traffic inside each ONU (intra-ONU).

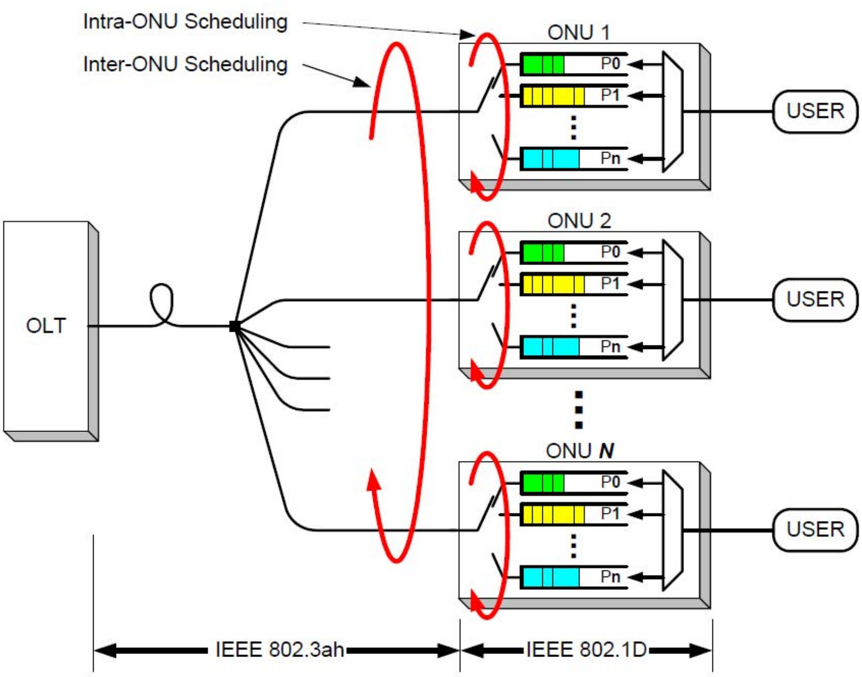


Fig. 3.2 The Scheduling mechanism intra/inter ONU (Source: [5])

In [127], the authors proposed an algorithm named *subMOS-IPACT* for a 10G-EPON system. This algorithm provides bandwidth guarantee with different granularity to individual ONUs, multi-ONU customers, and subgroups of ONUs. Meanwhile, it is able to provide bandwidth guarantee with pre-defined SLAs for both priority and no priority services. Therefore, the differentiated services within ONUs can be achieved.

Wireless local area networks: The distributed coordination function (DCF) in IEEE 802.11 standards was originally designed for best-effort service. In wireless network MAC, the transmission mechanism is designed as follows: If the transmission medium idles for a time period longer than a distributed inter frame space (DIFS), the host transmits the data packet. Otherwise, it enters the backoff process: A random value between 0 and CW (Contention Window) is calculated, and then this value is assigned as a backoff timer to defer data transmission. When the backoff time expires, the host starts to send the data to the medium. However, if there are two or more mobile hosts transmitting their data simultaneously, a collision occurs. In this case, an increased backoff time is calculated, and is set to the initial value of CW_{min} after every successful transmission.

To enable differentiated services in wireless LANs, Veres *et al.* [128] proposed to set different initial values for the contention window CW_{min} for different service classes. This ensures that a service class with a smaller content window has a higher probability to transmit its data first, no matter whether the collision is going to happen or not. Consequently, such setting of the contention window brings lower delays to the higher priority service classes.

3.2.3 Network Layer

The IP protocol suite is located at the network layer where the data packets are forwarded through the routing protocol. It is because of this forwarding mechanism that most protocols implementing differentiated services exist in the network layer.

Early Attempts

Before formal and feasible architectures were established, three early attempts had been proposed:

1. IP precedence: Type of Service (TOS)

2. Internet Stream protocol (ST)

3. Integrated service (IntServ)

However, these three attempts have been said to be "missing the mark" [129].

Type of Service (TOS):

This idea was considered in the original IP design document RFC 791 [130]. As shown in Fig. 3.3, the second byte of the IPv4 header was designed for Internet service quality selection. The first 3 bits in the TOS field are called "precedence", which indicates the priority of the traffic: priority '0' gets the lowest and priority '7' gets the highest. In a later document, RFC 1349 [131], the fourth to the sixth bits in the TOS field are used for specifying low delay, throughput, and reliability, respectively. And the seventh bit was defined as the "lowcost" field. The last bit of TOS is set to be '0' and named as "Must Be Zero" (MBZ). The TOS field is further modified for the Differentiated Services Code Point (DSCP) [6] in the DiffServ mechanism, and will be discussed later in this chapter. Although the concept of "type of service" was considered at the very beginning when the IP protocol is designed, the TOS field, however, has been generally ignored.

Internet Stream Protocol:

This protocol was first proposed by James W. Forgie in 1979 [132], and later revised in RFC 1190 [133] and RFC 1819 [134]. The Internet Stream Protocol (a.k.a. ST) was intended to provide QoS of Internet stream traffic. The idea of ST is to make real-time traffic travel in

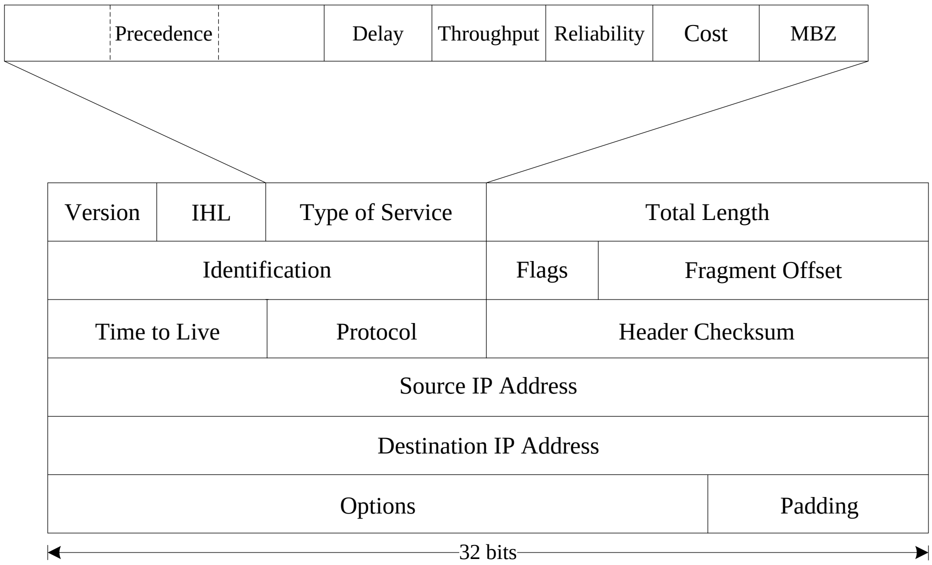


Fig. 3.3 The illustration of TOS

the network with other non-real time TCP/IP traffic in parallel through an add-on connection-oriented protocol. However, like TOS, the use of this approach is very limited except for experimental work.

The TOS field and ST have been rarely used. It is not feasible to add an 8-bits octet in an IP header or run a parallel protocol with TCP/IP protocol in a hierarchical network environment. An architecture or framework customized for differentiated services is needed.

The Integrated Service (IntServ) Architecture:

Researchers believed that the Internet's infrastructure must be modified to better support applications in a real-time environment. In 1994, IETF published Integrated Service (IntServ in short) in RFC 1663 [135]. The IntServ architecture aims to serve real-time traffic including guaranteed and predicted services. IntServ was designed to be embedded in the original best effort IP architecture, rather than made into a new system.

The basic principle of IntServ is to establish a virtual link between two end hosts. The IntServ architecture requires that every router in the system need to be IntServ compatible. To achieve the goal of implementing fine-grained QoS, the IntServ architecture also requires every application to do an individual reservation, which was described by "Flow specs". The Flow Spec includes two parts: Traffic specification "TSPEC" and Request specification "RSPEC". The reservation is made through the Resource reSerVation Protocol (RSVP) [136–140].

To clearly illustrate how IntServ works, Fig. 3.4 shows an example of setting up a communication process under the IntServ architecture. The red dot line represents the signaling path. All the intermediate routers (including edge routers: ER1, ER6, and core routers: CR1, CR4, CR3) between the sender and the receiver need to support the IntServ protocol to guarantee the QoS. During the communication/transmission, the resources occupied by this transmission can not be used by other requests.

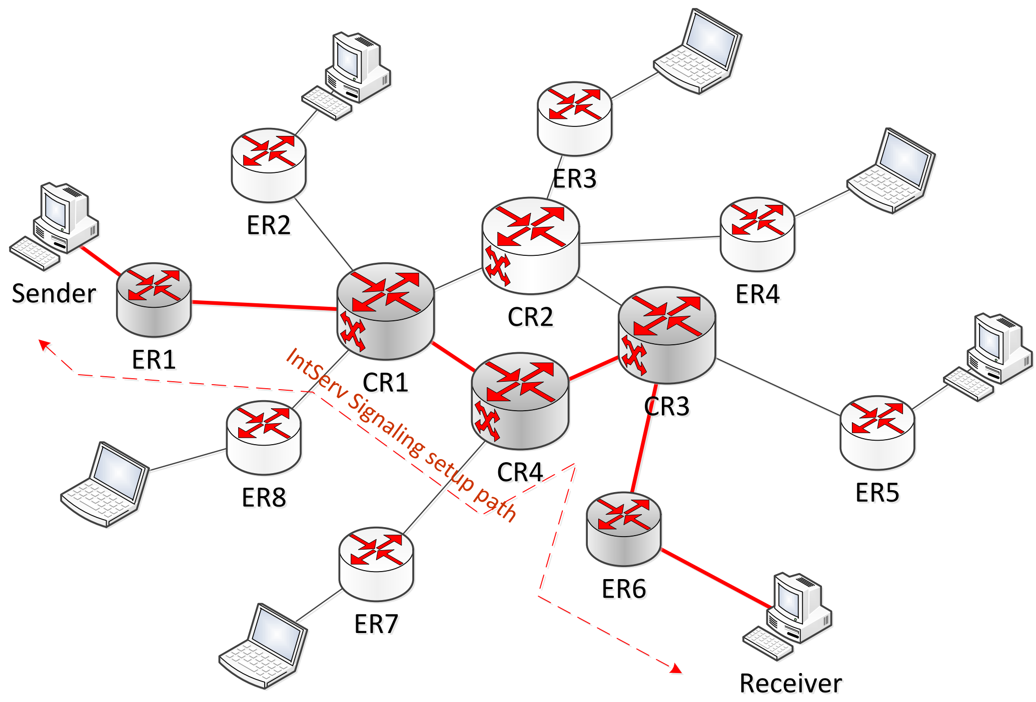


Fig. 3.4 An illustration of IntServ

In the IntServ architecture, a single connection between two end hosts is usually defined as a "flow", although, this definition is relaxed in [141, 142]. From a broader perspective, a "flow" can be detected from the 5-tuple of the IP header, namely, the transmission protocol of the data packet (such as TCP), the destination IP address, the source IP address, the port number of destination, and the port number of source. Also, the value of the DS field, which will be discussed next, can be used as a flow identifier. Such practices enlarge the granularity of the differentiated services entities, thereby increasing the feasibility of the architecture.

The Differentiated Services (DiffServ) Architecture:

IntServ stipulates too many details to efficiently allocate the resource to each flow, the feasibility of IntServ is relatively low. To overcome the shortcomings of IntServ, a coarse-grained service framework is needed.

IETF published RFC 2474 [6] and RFC 2475 [143] in December 1998. The former intended to provide scalable differentiated services, and the later defined the scalable architecture of implemented differentiated services (DiffServ in short) on the Internet, respectively. DiffServ is sometimes known as the “Soft” QoS model. The significant advantages of DiffServ are that it matches the Internet architecture well and that it can be deployed with very little modification, thus adding minimum complexity as needed [129]. Different from IntServ, DiffServ does not maintain the information of every flow, instead, it provides the quality guarantee in a coarse-grained manner. Specifically, DiffServ divides the traffic into different groups with different traffic demands, which are named as “Per Hop Behavior” (PHB).

In DiffServ, the implementation of differentiated services depends on the following two elements, namely PHB and codepoint.

PHB is a description of a class of data packets with the same QoS properties. It must be detailed enough to be compatible with the nodes of the differentiated services. The boundaries of a differentiated services area are also mentioned in [143]. Classifiers and traffic conditioners are deployed at the edge of the differentiated services areas. Generally speaking, an ingress router is an upstream node of differentiated services, and an egress router is a downstream node of differentiated services.

The codepoint is a special value that describes the PHB. As mentioned before, the DS field replaces the second octet "TOS" field in the IPv4 header, and replaces the "Class" octet in the IPv6 header as well. Fig. 3.5 shows the layout of the DS field. The first 6 bits form the Differentiated Service Code Point (DSCP); theoretically, there are total $2^6 = 64$ codepoints. The last two bits are marked as unused.

```
  0   1   2   3   4   5   6   7
+---+---+---+---+---+---+---+---+
|         DSCP          |  CU   |
+---+---+---+---+---+---+---+---+

  DSCP: differentiated services codepoint
  CU:   currently unused
```

Fig. 3.5 The layout of DSCP (source: [6])

Using the same topology in describing the IntServ, Fig. 3.6 gives an example of setting up a communication process using the DiffServ architecture. The blue solid line represents the forwarding path. All the classification information is collected at edge routers ER1 and ER2, and forwarded by core routers: CR1, CR4, and CR3.

A comparison between IntServ and DiffServ:

After introducing IntServ and DiffServ, the following table 3.1 helps to understand the advantages and disadvantages of the two architectures:

Given that DiffServ is more flexible, has less overhead, and easier deployment than IntServ, DiffServ has become the de facto standard [144] for differentiated services on the Internet. Based on the DiffServ architecture, Chaudhuri [145] proposed a QoS mechanism that can clearly improve the quality of real-time communications.

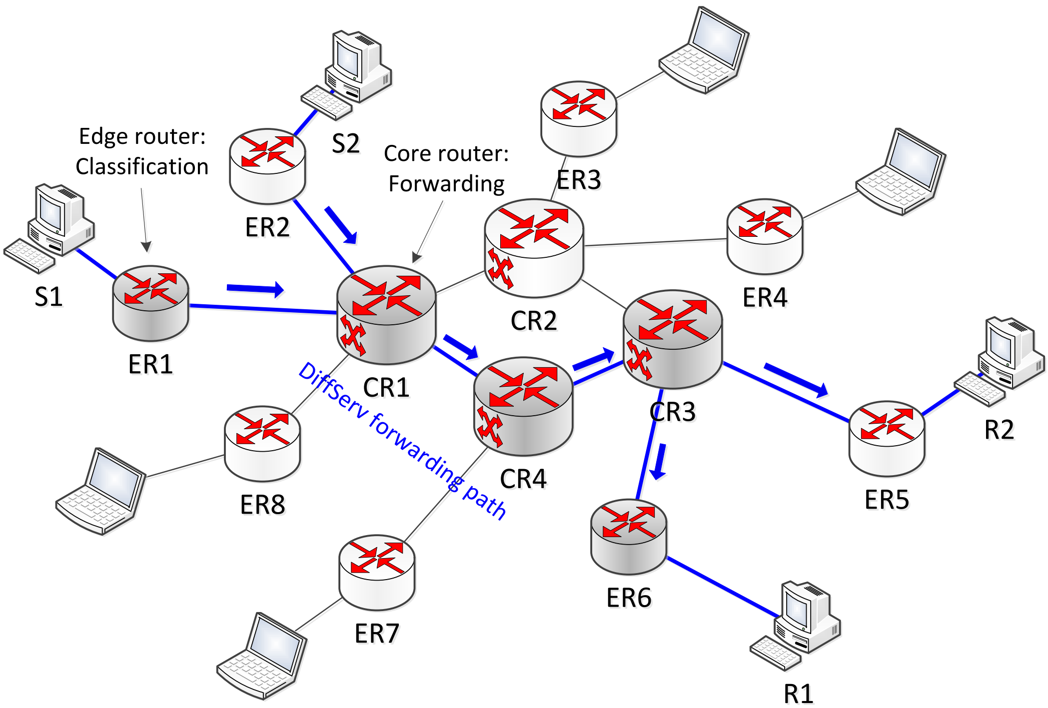


Fig. 3.6 An illustration of DiffServ

	Pros	Cons
IntServ	Per-flow QoS guaranteed; Suitable for managing flows in small networks;	Low scalability; High cost including flow signaling and memory of states; Difficult to operation and maintenance; Support limited traffic classes
DiffServ	High scalability; No reservation (protocol) needed; Easy to operation and maintenance; Support multilevel traffic classes;	Need to coordinate QoS across different DiffServ areas

Table 3.1 A comparison of IntServ and DiffServ

DiffServ over MPLS

We have discussed some attempts to implement differentiated services at the network layer. A technology named MPLS (Multi-Protocol Label Switching) adopts label switching instead of conventionally checking routing tables in the routers. It is regarded as located between the data link layer and the network layer (layer 2.5).

Based on MPLS, engineers further proposed an architecture to support differentiated services, which was originally published in the document RFC3270 [146]. After that, some hardware equipment manufacturers like Cisco [147, 148] and HUAWEI [149] also released various types of equipment (mainly routers) that support this architecture.

3.2.4 Transport Layer

TCP and UDP are located in layer 4 (transport layer) in the OSI 7 layers model. TCP is a connection-oriented protocol for reliable communications, while UDP is a connectionless protocol without guaranteeing the integrity of the data being transmitted. Protocols in the transport layer take charge of the sending data packets, for example, a TCP connection adjusts its sending rate (the size of the congestion window determines how many packets can be sent at a time) according to preset thresholds and the additive increase multiplicity decrease (AIMD) algorithm to avoid congestion. There are no widely accepted standards such as IntServ and DiffServ in the transport layer, so researchers have tried different ways to implement service differentiation in TCP and UDP. Most of them adjust the dropping probability of the incoming packets.

In order to enable ISPs to achieve service differentiation among different TCP flows, the use of weighted proportional fairness was studied in [150]. The weight of each TCP flow depends on the price being paid. In [151], a packet admission control mechanism named "RIO"(a modification of the well-known TCP/AQM mechanism "RED") was used to implement service differentiation. Its basic principle is to adjust the packet drop precedence of different queues to realize the priority. Using similar principles to improve the performance of TCP flows can be found in [152, 153]. Due to TCP and UDP serve different types of services, Lee *et al*. [154] introduced the Two Marking System (TMS) to fairly share the bandwidth to each flow for targeted differentiated performance (sending rate). The principle of marking system is the same as before: Set the dropping probability precedence of incoming data packets.

3.2.5 Session, Presentation and Application layer

In fact, we will not discuss the deployment of differentiated services in detail at each layer, since the protocols locate in these layers usually work interactively. For instance, the SIP protocol in the application layer (layer 7) cooperates with mechanisms in the session layer (layer 5) to establish a voice conversation between end-users, as shown in Fig. 3.1.

AS higher layers mean higher degrees of freedom, users and applications are given more flexibility and controllability of data traffic when using protocols (e.g., FTP and HTTP) located in higher layers. However, more freedom does not mean less controversy. An example can be found in [155], where the author proposed a mechanism, which gives end-users the right to prioritize the network traffic. A few comments gave positive remarks on this mechanism as it resolves the controversy around net neutrality for end-users have the right to choose the prioritization on which traffic, there is no one to blame. However, this approach is palliative, it does not eliminate network resource competition between users. Priority between a certain user's applications cannot resolve the high latency and service blocking caused by network congestion.

3.3 Summary

This chapter reviews the technical aspects of Internet Differential Services. Before the emerging of debate on network neutrality, engineers and researchers had foreseen the request of ever-growing data traffic demands. To meet such demands, they had established standards and proposed multiple ways to implement differentiated services on the Internet.

We discussed the implementation of differentiated services at various levels from the perspective of an OSI multi-layer model. Through a comprehensive analysis, it can be seen that standardized and systematic differentiated services are mainly located in the network layer and the transport layer. The distributed, high-scalability, low-overhead QoS mechanism (such as DiffServ) is more suitable for implementing differentiated services. Higher levels bring higher degrees of freedom, but additional degrees of freedom do not solve the problem of competition for network resources.

In the next chapter, we will use queueing theory to analyze the implementation of differentiated services in the last mile.

Chapter 4

Prioritized Services in the Last Mile

Net neutrality regulation restricts ISP to offer differentiated services to both CPs and end-users. However, differentiated service is prevalent in common business. As discussed in the previous chapter, differentiated services are feasible and are designed to be enabled in various architectures and network protocols. In this chapter, we study the impact of differentiated services from several perspectives. First, we study the impact of differentiated services through a series of simulations, from the perspectives of ISPs, and discuss the cost-saving effects. Second, we discuss the impact of differentiated services from the end-user perspective, using latency as the main metric. Third, we categorize traffic by service protocols, and investigate the impact of differentiated services based on real datasets.

4.1 Impact of Introducing Differentiated Services on ISPs

In this section, we study the impact of introducing differentiated services from the ISPs' perspective. Compared to the original one-tier service architecture, we first analyze the changes in QoS, and then study the benefits that ISPs can yield.

4.1.1 Analytical Tools

Call blocking probability theory (or Erlang formula) was proposed by the Danish mathematician Agner Krarup Erlang in 1909. It has been widely used in telephone networks in "hard blocking" mode for system capacity dimensioning and planning, and for call blocking probability analysis on the existing telephone network. The Erlang formula considers a scenario that when a channel is full, newly arrived calls are blocked. However, the Internet is a data network, which is different from telephone networks. Newly arrived users are allowed to enter and use the network resources with the existing users. Due to limited network resources (e.g., bandwidth), the concurrent users' Experience of Quality (QoE) is reduced, e.g., video quality degradation or extra buffering caused by increasing latency, and file download stall. In this case, the hard blocking probability of the Erlang formula cannot be directly applied to data networks. A recent white paper [156] has proposed a new method to convert the traditional hard blocking probability to "experience blocking" (EB), which is used to measure the QoS in mobile broadband networks.

Applying this new method to derive the so-called "experience blocking" has been planned for future work. Under all circumstances, the blocking probability in the Erlang formula correlates positively with the latency of the data network user. In this chapter, we build scenarios similar to actual data networks and use the Erlang B formula to assess the impact of differentiated services, and to study the impact on ISPs' costs. We try to convert the data network into an equivalent "M/G/k/k" (Erlang B) system. For example, a smooth high definition video streaming service on YouTube requires an average bandwidth of 5Mbps [157, 158]. Assuming a local ISP owns a total bandwidth of 10Gbps, then the number of equivalent channels is 10Gbps / 5Mbps = 2000.

4.1.2 Traffic Model: Loss System with Poisson Arrivals and Priorities (LSPP)

As in [159–161], we consider a loss system where the arrival of customers (requests) follows a Poisson process, and these customers have priority levels, so we call it a "loss system with Poisson arrivals and priorities" (LSPP). We use this system to model the traffic that goes through the last mile between end-users and ISPs. This is an extension of the queueing model known as M/G/*k*/*k* where different customers have different priorities.

The reason that we choose a model related to M/G/*k*/*k* as a modeling tool lies in that M/G/*k*/*k* has the property of insensitivity of the service distribution time. Specifically, the mean number of busy servers of the M/G/*k*/*k* and its blocking probability is insensitive to the shape of service time distribution. In addition to the arrival rate, we only need to know the first moment (i.e., the mean value) of the service time to get the blocking probability of the system. This feature provides us with great convenience and also improves the adaptability of the system as an analytical tool. We comment here that although LSPP is potentially likely to be insensitive which is demonstrated numerically, it was never rigorously proven to be insensitive.

The symbol "M/G/*k*/*k*" is according to Kendall's notation [162], where "M" stands for Markovian [163]. Accordingly, the first "M" implies that the arrival process is Poisson, and the time interval between two consecutive arrivals is exponentially distributed. The second "G" represents "General", which means the service time distribution is not specified. The first "*k*", in the third position, is the number of servers (or channels), and the second "*k*", in the fourth position, is the buffer space, namely, the maximum number of customers in the system. Since there is not extra buffer space (i.e., a loss system), any newly incoming customer will be blocked if all the servers (channels) are occupied.

4.1.3 Three levels of service (priority) setting

As discussed in Chapter 2, much of the research work on differentiated services on the Internet was inspired by the PMP scheme, which divides services into two classes: premium and ordinary. However, in practice, sellers often provide consumers with multiple levels of options, such as dividing airline tickets into first class, business class, and economy class on commercial flights. Generally, these products are divided into three levels: high, medium, and low. We believe that the three levels of service are more representative. In the following analysis, we divided the service into three classes (or priorities): the first-class represents the premium service, which has the highest price but is provided with the highest Internet speed; the second class represents the ordinary class which is affordable by most of the consumers; the third class, can be regarded as the government-funded free Internet access (e.g., free Wi-Fi), which has the lowest Internet speed.

4.1.4 Traffic Assumptions

The traffic assumption is set as follows: The arrival rate is 4000 requests per second. Here the request can be understood as a data burst, and the system can be regarded as service channel(s) of a local ISP. The transmission capacity of each link/channel is 10 Gb/s (i.e., 10×10^9 bit/s). According to [158], we set the packet size as 1400 Bytes, i.e., $1400 \times 8 = 11200$ bits, which is close to the average packet size of transmitting YouTube traffic. The average number of packets in a request is 800. Given the increasing proportion of video streaming on the Internet [157], we set the proportion of each priority traffic as follows: Priority 1 takes 40%, Priority 2 takes 50% and Priority 3 takes 10%. Then the overall offer load A can be calculated as: $4000 \times 1400 \times 8 \times 800/10^{10} = 3.584$ Erlangs. Accordingly, Priority 1 provides 1.4336 Erlangs, Priority 2 provides 1.792 Erlangs, and Priority 3 provides 0.358 Erlangs.

Generally speaking, the bandwidth of a broadband network is two-way, the ISPs send and receive data from users. However, judging from the application requirements of current broadband networks, the downlink demand is much higher than the uplink demand, that is, end-users often cannot get the desired bandwidth and cause congestion. Therefore, we will focus on the analysis of downstream traffic.

4.1.5 Wire-line Network Case

We begin with the wire-line network case, more specifically a single-node case that represents a bottleneck connection between an ISP and its subscribers. The topology is shown in Fig. 4.1. Such a topology also describes the one-side market, where the CPs can be considered as end-users. The ISP is located between its subscribers (end-users) and the core network. The single-node here refers to a single ISP that establishes a connection between end-users and the core network. Generally, end-users subscribe to only one ISP in order to obtain their Internet access service.

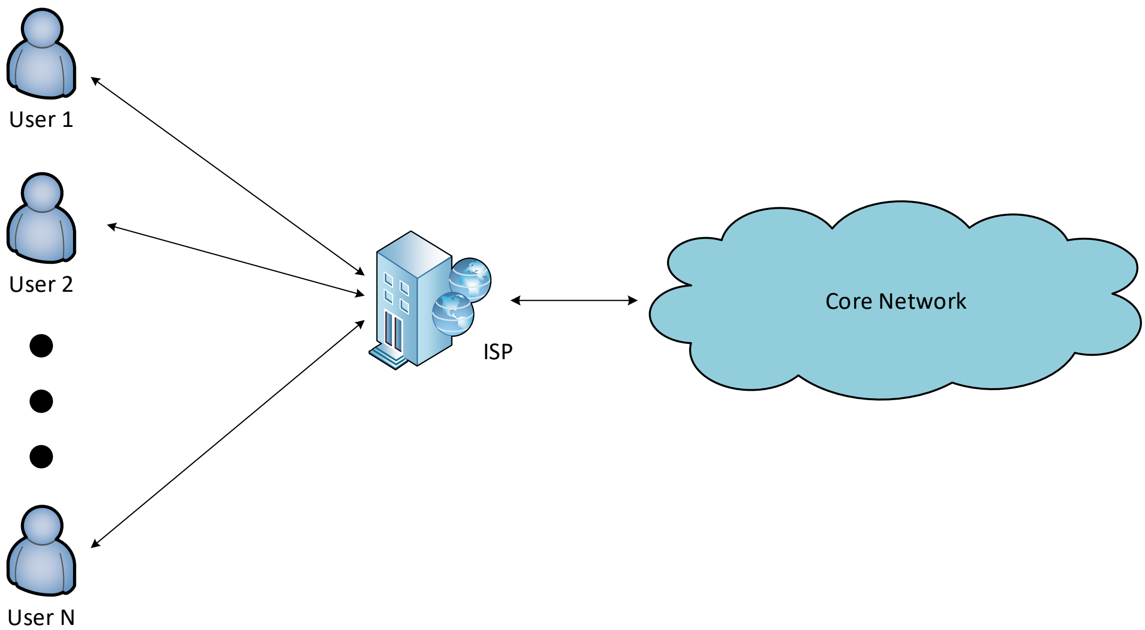


Fig. 4.1 The topology of single node (ISP) case

The QoS is measured in terms of hard blocking probability. As mentioned before, the Quality of Experience (QoE) of data networks is different from the QoS in circuit-switched networks (e.g., telephone network). When the number of concurrent users in the data networks increases, the QoE decreases in the form of increased service delay and buffering. In the following, the "hard blocking" probability calculated by Erlang B formula or derived by simulation can be regarded as the "soft blocking" of the service, which means extra delay. The higher the blocking probability, the higher the delay end-user experience.

In this model, we assume there are three priorities: $P1$ has the highest priority, $P2$ is in the middle and $P3$ is the lowest priority. How the mechanism of such differentiated service works is now explained: when a new request with priority P enters the system, it chooses a random channel if any available channel exists, otherwise, it checks whether there is an existing request being served with priority lower than P, if true, the new coming request will preempt a random selected existing request being served, which means the preempted request is blocked. Otherwise, the new incoming request is blocked. This mechanism ensures that if there is no congestion in the network, all the service requirements can be served regardless of their service level, and when congestion occurs, the higher priority requests preempt existing requests with lower priority.

Analytical Calculation

Before conducting the simulation, the analytical expression is given in the following:

$$\begin{aligned} &B_1 = E(A_1, C), \\ &B_2 = (E(A_1 + A_2, C) \times (A_1 + A_2) - B_1 \times A_1)/A_2, \\ &B_3 = (E(A_1 + A_2 + A_3, C) \times \\ &(A_1 + A_2 + A_3) - B_1 \times A_1 - B_2 \times A_2)/A_3 \end{aligned} \tag{4.1}$$

where function $E(*)$ stands for Erlang B formula, and B_1, B_2 and B_3 represent the blocking probability of priority $P1$, priority $P2$ and priority $P3$ traffic, respectively. Parameters A_1, A_2, and A_3 are the corresponding offered load of each priority.

Table 4.1 gives the theoretical calculation results of the blocking probability of the single-node case.

Table 4.1 The results of analytical calculation of a single node (ISP) case

	Blocking Probability			
Number of channel	Priority 1	Priority 2	Priority 3	Average
1	0.5890861	0.9027561	0.9483742	0.8134055
2	0.2968922	0.7557189	0.8690358	0.6405490
3	0.1242473	0.5708650	0.7563220	0.4838114
4	0.0426318	0.3815749	0.6118927	0.3453665
5	0.0120758	0.2237323	0.4496838	0.2284973
6	0.0028770	0.1149941	0.2941581	0.1373431
7	0.0005889	0.0520031	0.1690623	0.0738848
8	0.0001055	0.0208274	0.0850582	0.0353304
9	0.0000168	0.0074503	0.0376253	0.0150308
10	0.0000024	0.0024024	0.0147665	0.0057238
11	0.0000003	0.0007045	0.0051967	0.0019672
12	0.0000000	0.0001894	0.0016567	0.0006154
13	0.0000000	0.0000470	0.0004828	0.0001766
14	0.0000000	0.0000108	0.0001296	0.0000468
15	0.0000000	0.0000023	0.0000323	0.0000115

Fig. 4.2 shows the theoretical calculation results. This figure serves as a reference for subsequent simulation results. Intuitively, the higher priority traffic suffers lower blocking probability, users with higher priority suffer the lowest service delay and enjoy the highest QoE.

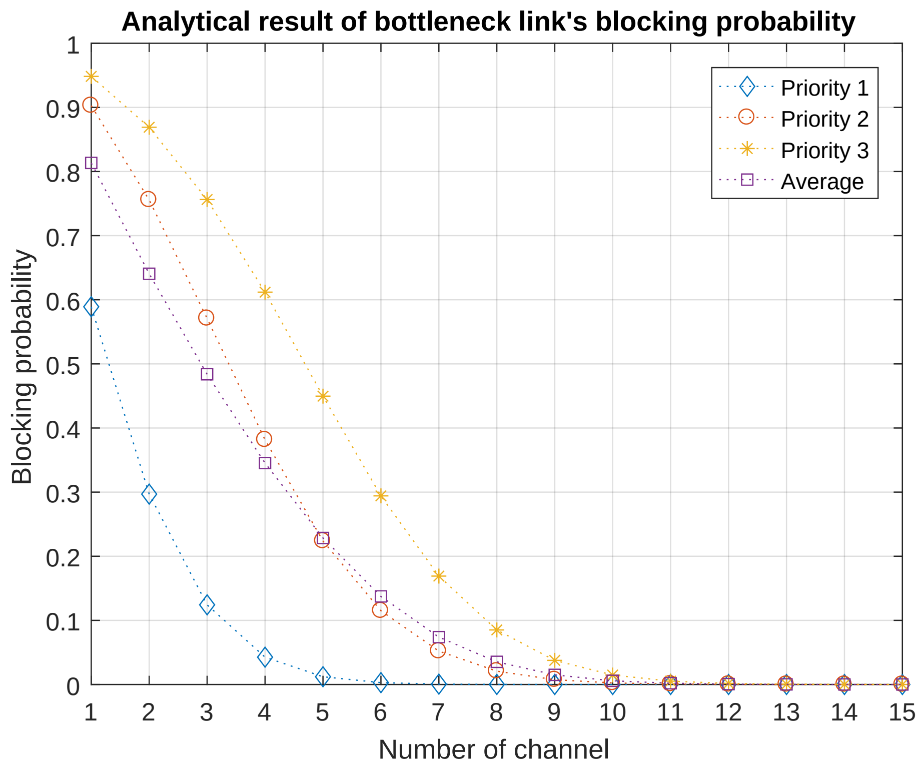


Fig. 4.2 The blocking probability results of analytical calculation of a single node (ISP) case

Markov Chain Simulation

In the simulation part, we begin with a Markov Chain simulation. We keep the same parameters as described above, and the proportion of the offered load of three different priorities are given by 50%, 40%, and 10%, respectively. In all the following simulations, the total number of arrivals is set to 1×10^6, and we run the simulation 6 times to get a 95% confidence interval.

Table 4.2 provides the Markov Chain results of the blocking probability of the single-node case.

Table 4.2 Markov chain simulation results

	Blocking Probability			
Number of channel	Priority 1	Priority 2	Priority 3	Average
1	0.5897121	0.9034156	0.9493660	0.8141646
2	0.2965920	0.7575718	0.8622218	0.6387952
3	0.1239682	0.5732208	0.7364617	0.4778836
4	0.0415752	0.3852950	0.5832059	0.3366920
5	0.0122406	0.2309958	0.4062937	0.2165100
6	0.0028708	0.1231691	0.2603330	0.1287910
7	0.0005485	0.0575357	0.1447198	0.0676014
8	0.0000794	0.0241923	0.0683603	0.0308773
9	0.0000050	0.0087347	0.0298038	0.0128479
10	0.0000000	0.0029342	0.0120494	0.0049945
11	0.0000000	0.0008988	0.0037347	0.0015445
12	0.0000000	0.0002080	0.0010348	0.0004143
13	0.0000000	0.0000599	0.0003215	0.0001272
14	0.0000000	0.0000240	0.0000404	0.0000214
15	0.0000000	0.0000120	0.0000000	0.0000040

Fig. 4.3 shows the Markov chain simulation result of a single node case.

Discrete Event Simulation

Using the same traffic assumptions, we conduct discrete event simulations.

Fig. 4.4 shows the logic of priority preemption and the process of the discrete event simulation.

Table 4.3 and Fig. 4.5 provide the results of the discrete event simulation.

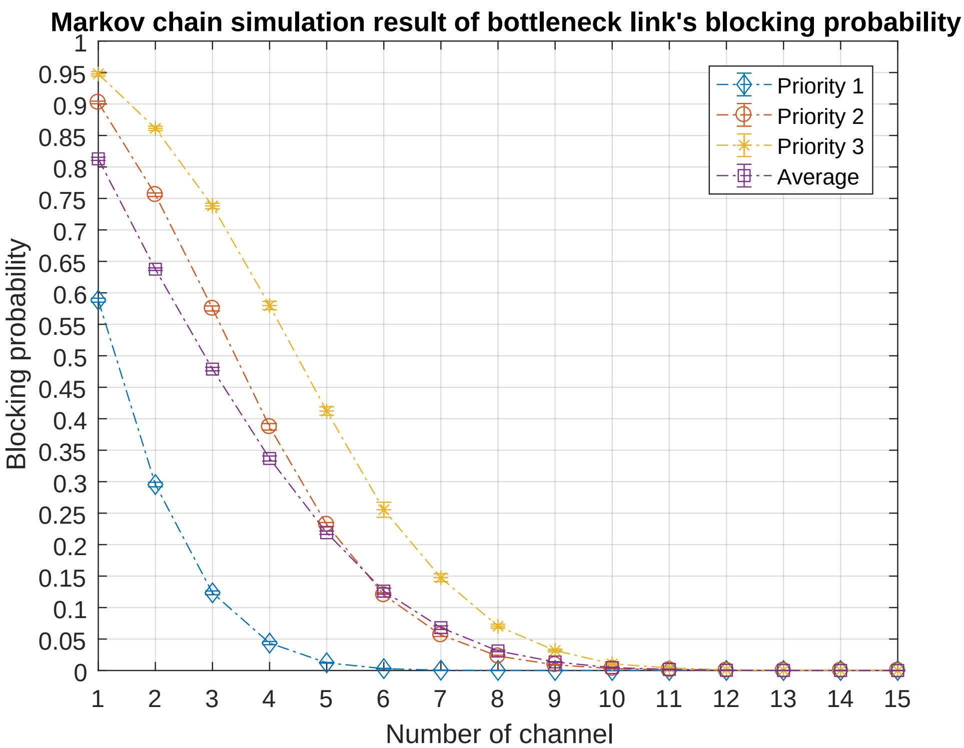


Fig. 4.3 Markov chain simulation blocking probability results of a single node (ISP) case

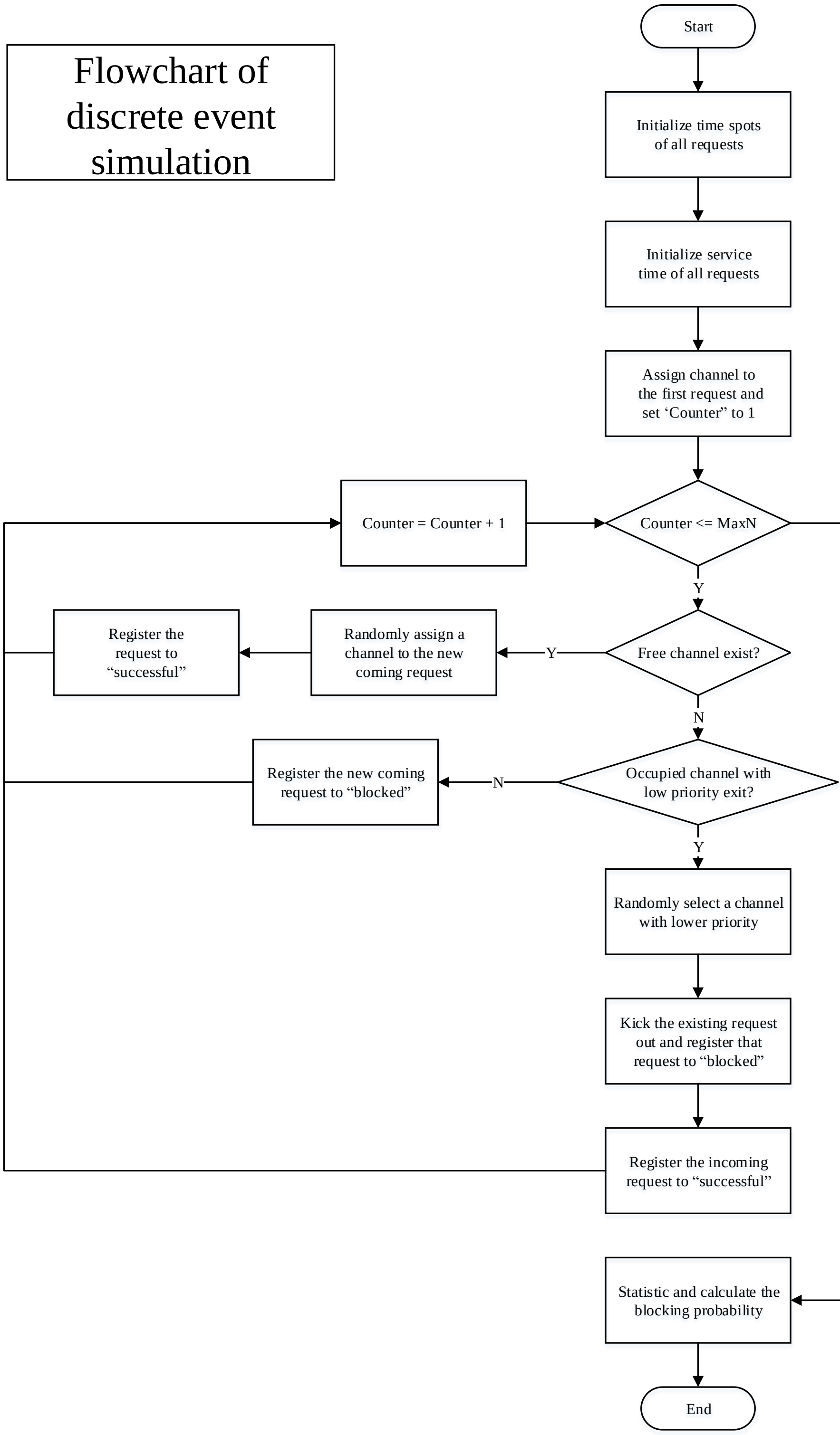


Fig. 4.4 The flowchart of the discrete event simulation

Table 4.3 Discrete event simulation results

	Blocking Probability			
Number of channel	Priority 1	Priority 2	Priority 3	Average
1	0.5891126	0.9035102	0.9491237	0.8139155
2	0.2960460	0.7560774	0.8603799	0.6375011
3	0.1245678	0.5737640	0.7346984	0.4776767
4	0.0426448	0.3879197	0.5818907	0.3374851
5	0.0112794	0.2300632	0.4098787	0.2170738
6	0.0027609	0.1212673	0.2593712	0.1277998
7	0.0004929	0.0567183	0.1439625	0.0670579
8	0.0000825	0.0234883	0.0708562	0.0314757
9	0.0000025	0.0089109	0.0294006	0.0127713
10	0.0000000	0.0029586	0.0108876	0.0046154
11	0.0000000	0.0008795	0.0035208	0.0014668
12	0.0000000	0.0002599	0.0012910	0.0005170
13	0.0000000	0.0000640	0.0004305	0.0001648
14	0.0000000	0.0000180	0.0001399	0.0000526
15	0.0000000	0.0000140	0.0000700	0.0000280

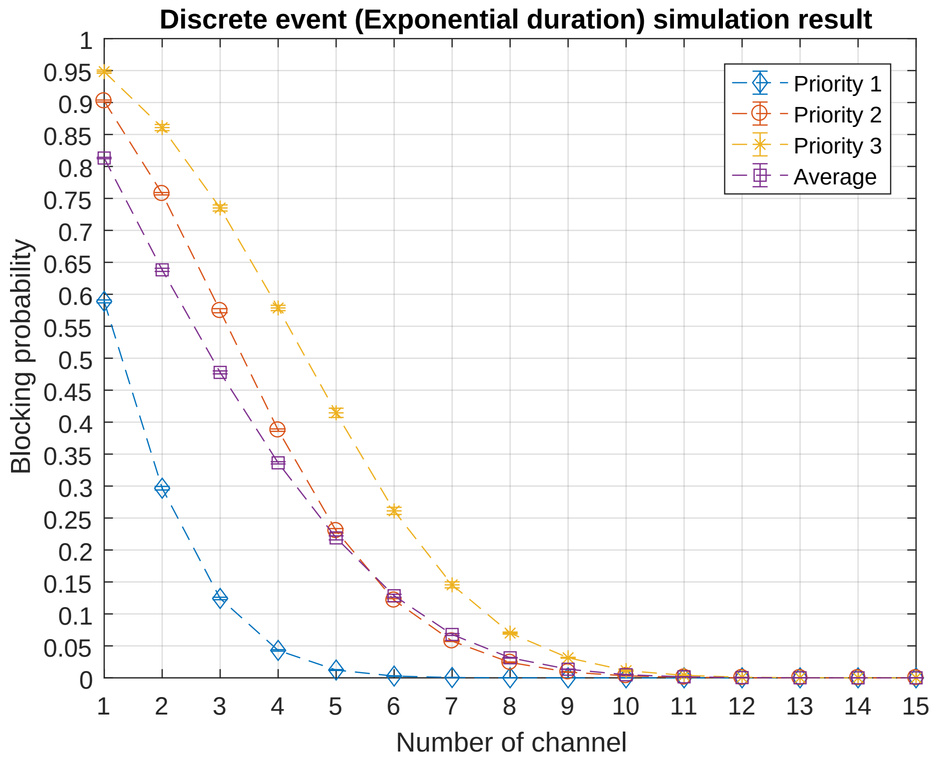


Fig. 4.5 Discrete event simulation blocking probability results of a single node (ISP) case

From Figures 4.2, 4.3, and 4.5, it can be seen that the two kinds of simulation results are consistent with the analytical calculations.

Discussion of the numerical results

By comparing theoretical calculations with experimental results, we see that the experimental results are highly consistent with the theoretical calculations. A different QoS is obtained through differentiated services.

For ease of reading, we take the 5% blocking probability as a reference value. We also plot the average blocking probability under a single service mechanism as a reference. From Fig. 4.2, Fig. 4.3 and Fig. 4.5, we find that introducing differentiated services can indeed save costs for ISPs. Based on the traffic assumptions, under the single service mechanism (purple curve with square data points), the ISP needs 8 channels to meet the requirement that blocking probability of all traffic is under 5%. Under the differentiated services mechanism, the highest priority (blue curve with diamond data points) can meet the requirement with only 4 channels, which means that only 4 channels are needed to satisfy the users with the highest delay requirements. Besides, if two more channels are added, the blocking probability of the second priority (red curve with circle data points) will be lower than the blocking probability of the single service mechanism (purple curve with square data points), which is undoubtedly a benefit for the second 50% end-user group.

Discrete Simulation with Pareto distributed Service times

We have used exponential distribution to simulate service duration, however, Internet traffic exhibits the characteristics of heavy-tailed distributed and self-similarity [164, 165].

For a certain range of its parameters has an infinite variance, the Pareto distribution possesses a heavy-tailed distribution [166, 167]. Therefore, it is regarded as a more accurate mathematical model of burst duration on the Internet. Considering that we adopted the Poisson process to simulate the arrival process of the requests, now the traffic can be modeled as a PPBP (Poisson Pareto burst process) [168, 169], where the arrivals follow the Poisson process and traffic duration follows Pareto distribution.

To model heavy-tailed burst-lengths, the Pareto distribution that we use has infinite variance but finite mean. This allows us to model the significant long bursts that characterize long-range dependent (LRD) traffic [170, 171] to occur in the model.

The complementary distribution function Pareto random variable distribution is given by:

$$Pr(d > x) = \begin{cases} (\frac{x}{\delta})^{-\gamma}, & x \geq \delta, \\ 1, & otherwise. \end{cases} \tag{4.2}$$

To model long-range dependent traffic, we keep the same average value of the service duration (800 packets) but set the variance of the service duration time to infinity. Table 4.4 shows the numerical results of the blocking probability and we plot them in Fig. 4.6.

Compared to the previous results (both analytical and simulations), we see that the blocking probabilities under Pareto distributed service time is slightly smaller than the previous ones.

Table 4.4 Discrete event simulation (Pareto distributed service time with infinite variance) results of a single node (ISP) case

	Blocking Probability			
Number of channel	Priority 1	Priority 2	Priority 3	Average
1	0.7164829	0.7164829	0.7164829	0.7164829
2	0.4243227	0.4243227	0.4243227	0.4243227
3	0.2421342	0.2421342	0.2421342	0.2421342
4	0.1201868	0.1201868	0.1201868	0.1201868
5	0.0504246	0.0504246	0.0504246	0.0504246
6	0.0238254	0.0238254	0.0238254	0.0238254
7	0.0087102	0.0087102	0.0087102	0.0087102
8	0.0037117	0.0037117	0.0037117	0.0037117
9	0.0011557	0.0011557	0.0011557	0.0011557
10	0.0004002	0.0004002	0.0004002	0.0004002
11	0.0001066	0.0001066	0.0001066	0.0001066
12	0.0000367	0.0000367	0.0000367	0.0000367
13	0.0000052	0.0000052	0.0000052	0.0000052
14	0.0000013	0.0000013	0.0000013	0.0000013
15	0.0000000	0.0000000	0.0000000	0.0000000

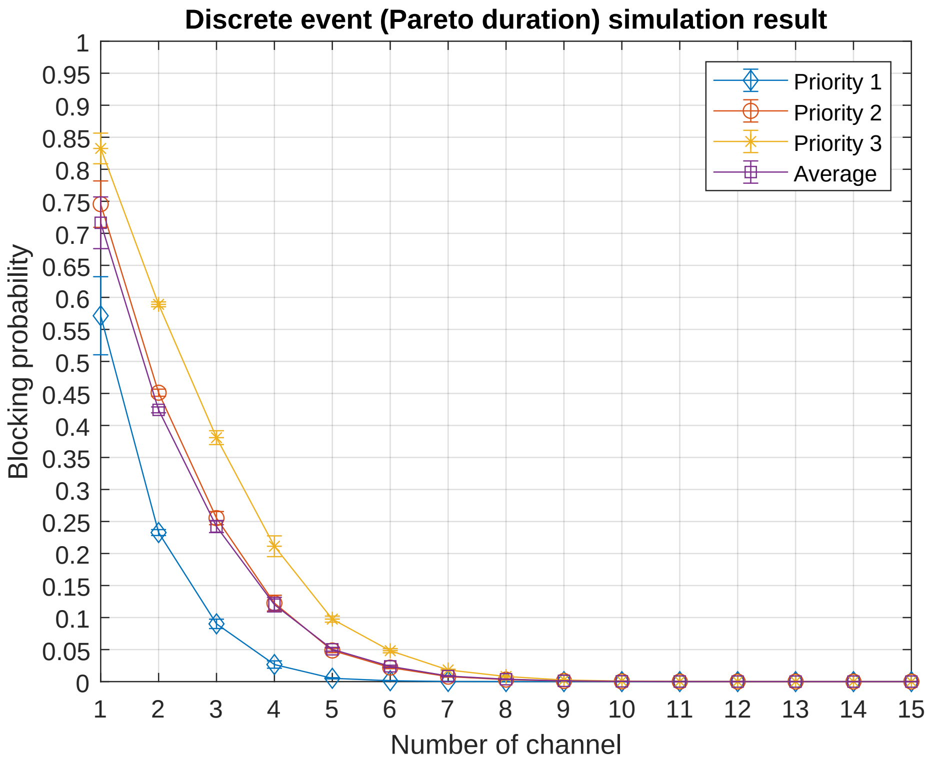


Fig. 4.6 Discrete event simulation blocking probability results of a single node (ISP) for the case of Pareto distributed service time with infinite variance

To find the reason, we checked the average number of data packets of each arrival (request) and found that in this set of simulations, the average number of data packets was 672.5134, which was less than the target value of 800. This is a well-known effect of generation of random deviates from a heavy-tailed distribution. The smallest value of the packets is 50, and the largest value of the packets is 73,787,053. We plot the histogram of the random deviates from the Pareto distribution with infinite variance in Fig. 4.7. In order to enable the requests with largest data packets to be clearly displayed on the Y-axis, we adopted the logarithmic scale for the Y-axis. It can be seen that most requests have a small number of data packets while a very small number of requests contain a large number of packets.

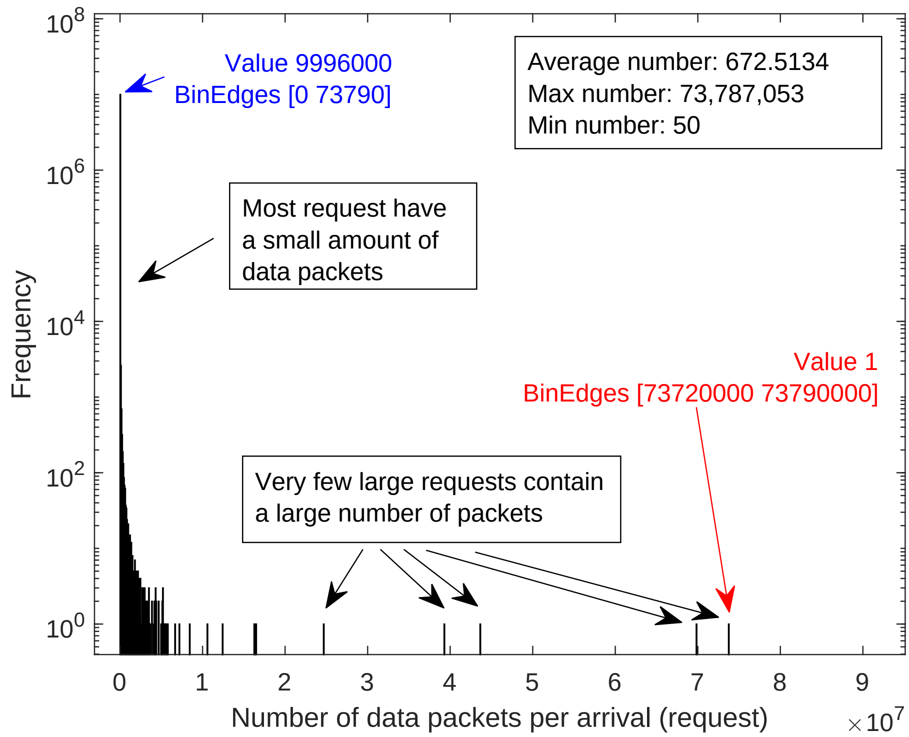


Fig. 4.7 The histogram of random deviates from Pareto distribution with infinite variance

When the variance of the Pareto distribution is infinite, in order to make the average of Pareto random deviates reach a value close to the target value, it is necessary to generate very large samples. However, it is computationally prohibitive to generate a very large sample of Pareto random deviates. If we want to let the average number of Pareto random deviates reach a value close to the target value (800), one way is to generate random deviates from the Pareto distribution with finite variance.

Now we use Pareto distribution with finite variance to model the service duration time, and conduct the simulation again. Fig. 4.8 shows the histogram of the random deviates' histogram, and Table 4.5 and Fig. 4.9 show the blocking probability results.

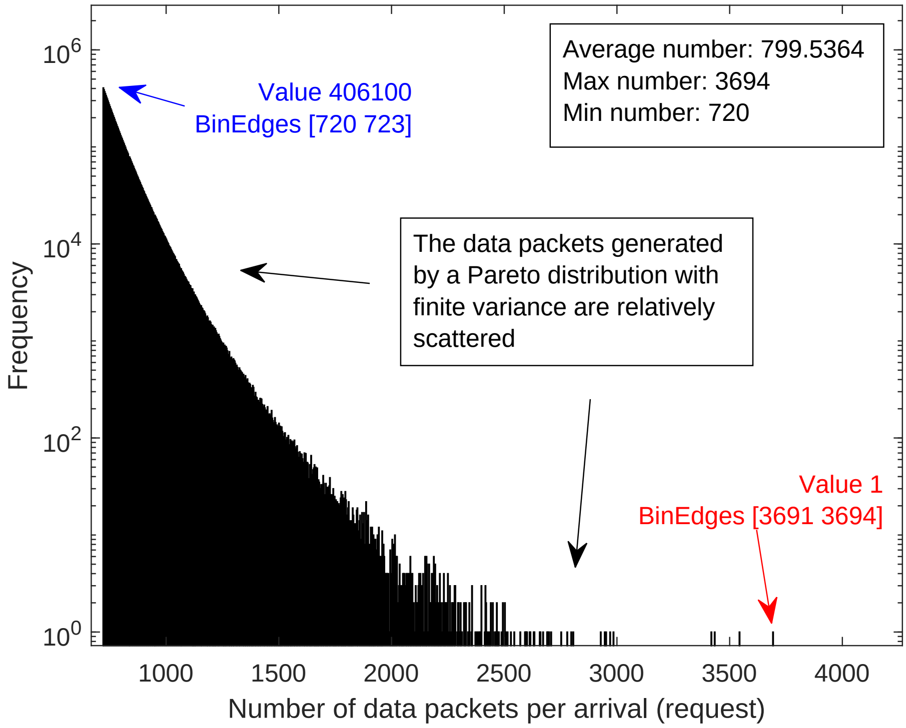


Fig. 4.8 The histogram of random deviates from Pareto distribution with finite variance

Table 4.5 Discrete event simulation (Pareto distributed service time with finite variance) results of a single node (ISP) case

	Blocking Probability			
Number of channel	Priority 1	Priority 2	Priority 3	Average
1	0.5888800	0.9169330	0.9678816	0.8245649
2	0.2962764	0.7416508	0.8732134	0.6370469
3	0.1238952	0.5174732	0.7075011	0.4496232
4	0.0424791	0.3139000	0.5047526	0.2870439
5	0.0122839	0.1671783	0.3149642	0.1648088
6	0.0027664	0.0817166	0.1735050	0.0859960
7	0.0006089	0.0363096	0.0846656	0.0405281
8	0.0001145	0.0151149	0.0369023	0.0173772
9	0.0000108	0.0055968	0.0144636	0.0066904
10	0.0000033	0.0020922	0.0052208	0.0024388
11	0.0000000	0.0006328	0.0016316	0.0007548
12	0.0000000	0.0001882	0.0004795	0.0002226
13	0.0000000	0.0000450	0.0001447	0.0000633
14	0.0000000	0.0000133	0.0000424	0.0000185
15	0.0000000	0.0000100	0.0000077	0.0000059

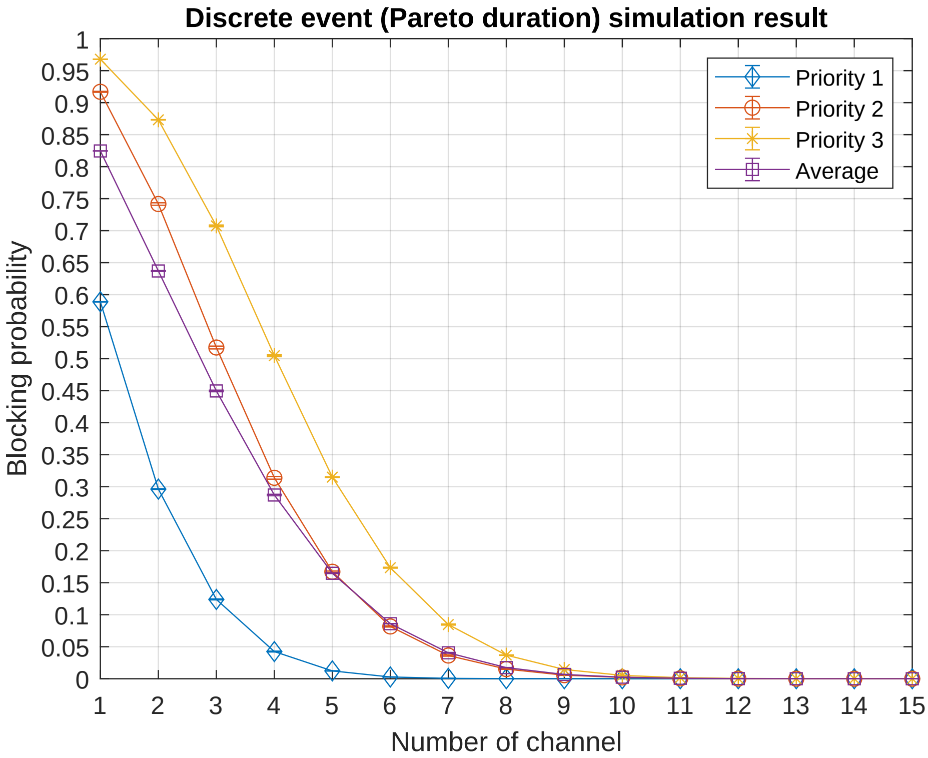


Fig. 4.9 Discrete event simulation results of a single node (ISP) for the case of Pareto distributed service time with finite variance

Comparing Fig. 4.7 and Fig. 4.8, we can see that the shape of the histogram has been changed, the data packets generated from a Pareto distribution with finite variance are relatively scattered, and the maximum value (3694) in Fig. 4.8 is much smaller than the maximum value (73,787,053) in Fig. 4.7. Meanwhile, the average number of the Pareto random deviates in Fig. 4.8 is 799.5364, which is very close to the target value of 800. More importantly, from Table 4.5 and Fig. 4.9, we find that the blocking probability of three priorities is very close to the theoretical calculation and previous simulations. This verifies that such LSPP system is insensitive to the service time distribution.

4.1.6 Wireless Network Case

Given that mobile devices are contributing more and more Internet traffic, it is necessary to simulate wireless network scenarios. To do so, we extend the prioritized traffic model into a mobile network case. Two wireless network cases will be tested, namely the regular seven cells case and the random 12 cells case.

A regular 7-Cell cellular network case

First, we use a regular seven-cell topology to illustrate the mobile network; the topology is shown in Fig. 4.10:

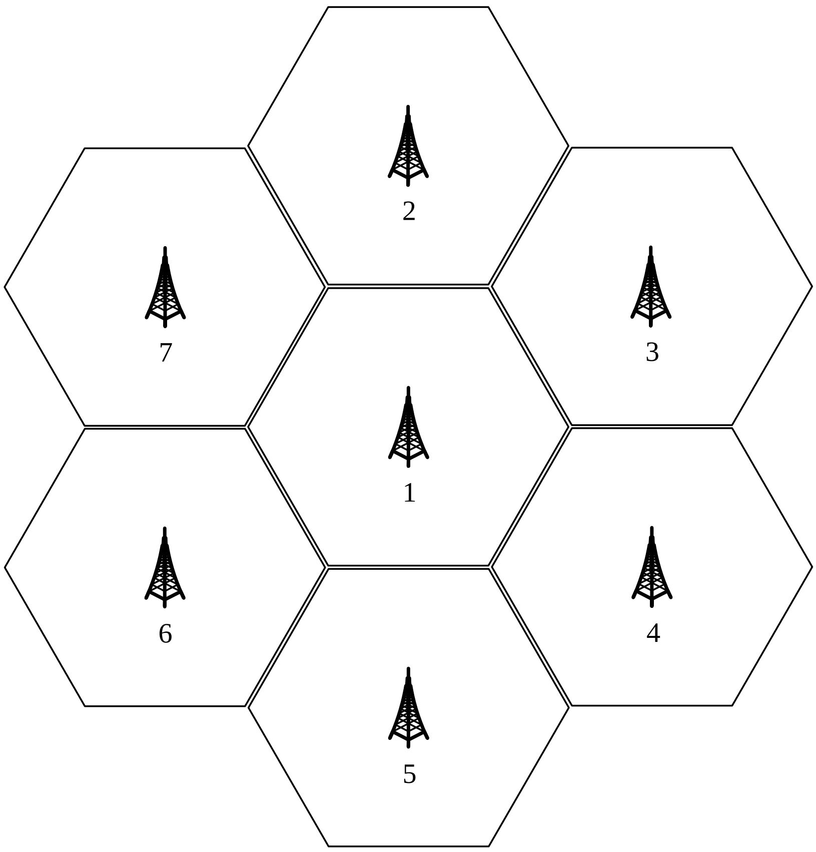


Fig. 4.10 The topology of the a regular 7 cell wireless network case

The seven-cell topology is a typical topology of the cellular system adopted both in academic research and industrial application. Each cell is represented by a standard hexagon.

A base station is located in the center of each cell, in order to maximize signal coverage. Seven cells consist of a cluster in a cellular system, where the frequency reuse factor is $1/7$.

Here we follow the same arrival process in the wireline network, namely the arrival rate λ. However, the arriving customers are now distributed evenly in the seven cells, i.e., the total arrival rate in each cell is one-seventh of the wire-line case. Besides, the transmission speed of each channel is set to 5 Gbps (5×10^9 bit/s). Compared with the transmission rate of the current cellular network, this setting is relatively high, but this may be in line with the upcoming 5G era. It is worth noting that there has been much work discussing the use of network slicing technique [172–178] to achieve differentiated services in 5G networks. Due to mobility nature, an end-user traveling from one cell to another cell is called "handover". The following matrix describes the handoff probability from one cell i to another cell j:

$$P(i,j)_{regular} = \begin{bmatrix} 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 0 & 1/3 & 0 \end{bmatrix}$$

where $P(i,j)$ is the probability when the handover happens, a customer who originates from cell i, moves to adjacent cell j.

Table 4.6 and Fig. 4.11 show the theoretical calculation of the blocking probability of each priority.

	Blocking Probability			
Channel number	Priority 1	Priority 2	Priority 3	Average
1	0.1699867	0.4318091	0.5472511	0.3830156
2	0.0171088	0.1082701	0.1879395	0.1044395
3	0.0011666	0.0176064	0.0416181	0.0201304
4	0.0000597	0.0020855	0.0064968	0.0028807
5	0.0000024	0.0001946	0.0007743	0.0003238
6	0.0000001	0.0000150	0.0000745	0.0000299
7	0.0000000	0.0000010	0.0000060	0.0000023
8	0.0000000	0.0000001	0.0000004	0.0000002
9	0.0000000	0.0000000	0.0000000	0.0000000
10	0.0000000	0.0000000	0.0000000	0.0000000

Table 4.6 The theoretical calculation of a 7 cells regular wireless network case

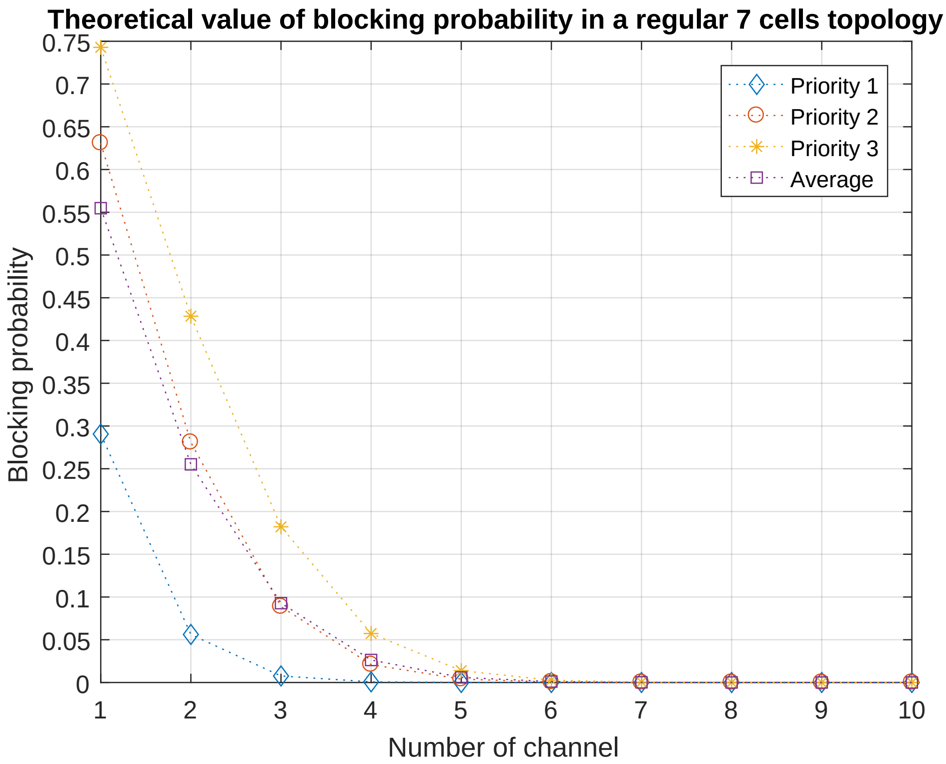


Fig. 4.11 The theoretical calculation of blocking probability of a 7 cells regular wireless network case

Next, we conduct a discrete event simulation. Table 4.7 and Fig. 4.12 show the results of the simulation.

Table 4.7 The discrete simulation of a 7 cells regular wireless network case

	Blocking Probability			
Channel number	Priority 1	Priority 2	Priority 3	Average
1	0.1859742	0.4561102	0.5617601	0.3590400
2	0.0239360	0.1418965	0.2215864	0.1029100
3	0.0024312	0.0337830	0.0638319	0.0243200
4	0.0001003	0.0066845	0.0159826	0.0050000
5	0.0000000	0.0011208	0.0034530	0.0009100
6	0.0000000	0.0001001	0.0007893	0.0001300
7	0.0000000	0.0000000	0.0002960	0.0000300
8	0.0000000	0.0000000	0.0000000	0.0000000
9	0.0000000	0.0000000	0.0000000	0.0000000
10	0.0000000	0.0000000	0.0000000	0.0000000

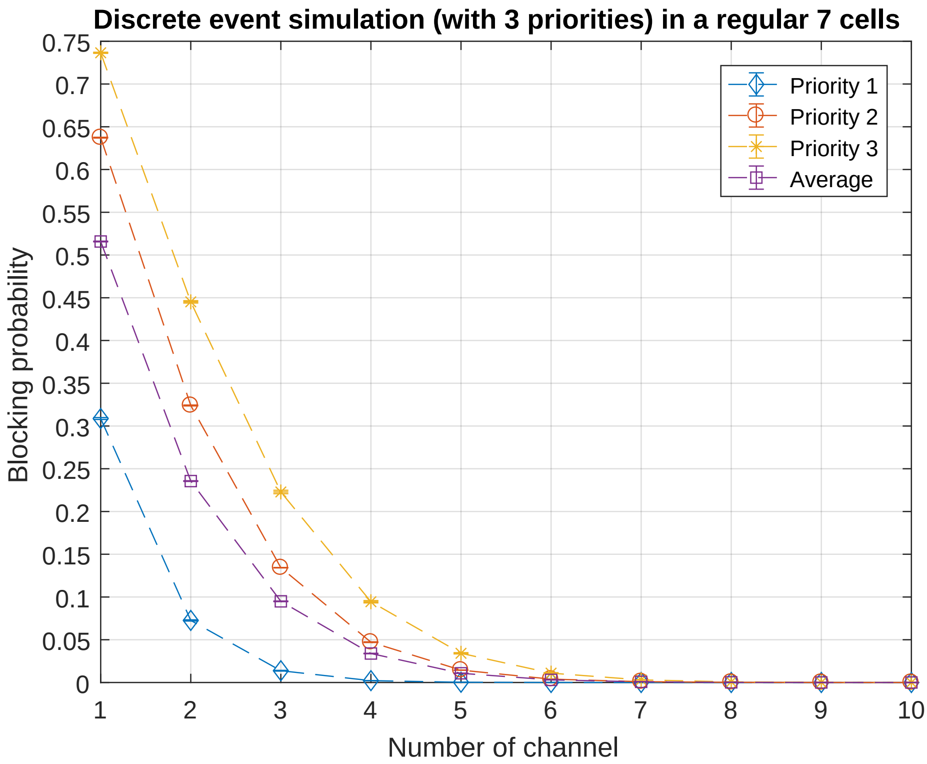


Fig. 4.12 The discrete simulation blocking probability results of 7 cells regular wireless network case

Random Topology

In reality, the distribution of mobile phone base stations is often irregular. We now test a cellular network in a random manner. An area is divided into 12 subregions. The topology of this random topology is shown in Fig. 4.13.

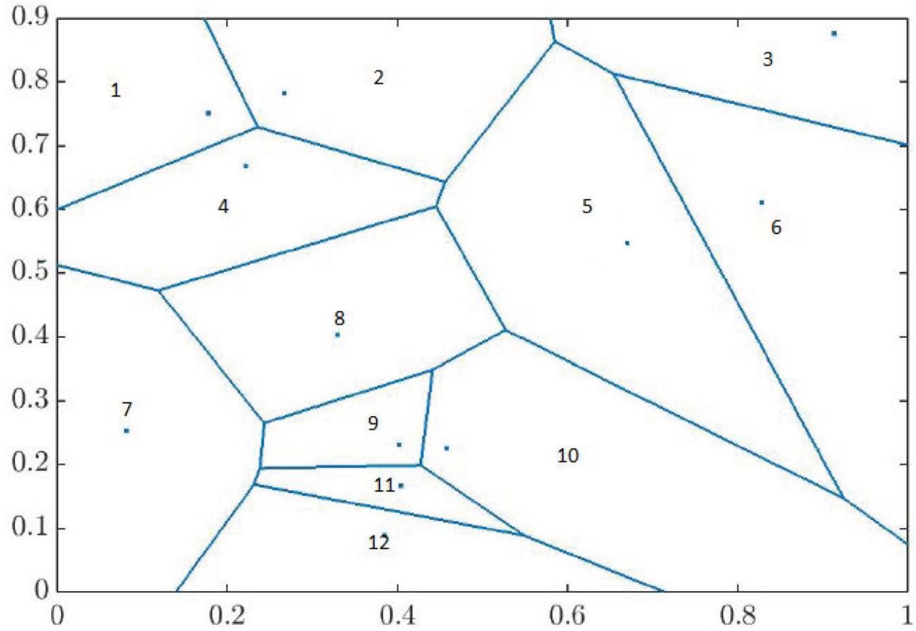


Fig. 4.13 The topology of a 12 cells random cellular network

We assume that a request in a cell has a uniformly distributed probability to move to another cell. The following matrix describes the probability that a request (call) move from cell i to cell j in this random topology:

$$P(i,j)_{random} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/5 & 1/5 & 0 & 0 & 1/5 & 0 & 1/5 & 1/5 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/6 & 1/6 & 0 & 1/6 & 0 & 1/6 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 0 & 0 & 0 & 1/5 & 1/5 & 0 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/5 & 1/5 & 0 & 1/5 & 0 & 1/5 & 1/5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 1/6 & 1/6 & 0 & 1/6 & 1/6 & 0 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \end{bmatrix}$$

Table 4.8 The theoretical calculation of a 12 cells random wireless network case

	Blocking Probability			
Channel number	Priority 1	Priority 2	Priority 3	Average
1	0.1928541	0.4750612	0.5928438	0.4202530
2	0.0225208	0.1366183	0.2314990	0.1302127
3	0.0017904	0.0258581	0.0596662	0.0291049
4	0.0001069	0.0035748	0.0108996	0.0048605
5	0.0000051	0.0003894	0.0015200	0.0006382
6	0.0000002	0.0000351	0.0001709	0.0000687
7	0.0000000	0.0000027	0.0000161	0.0000063
8	0.0000000	0.0000002	0.0000013	0.0000005
9	0.0000000	0.0000000	0.0000001	0.0000000
10	0.0000000	0.0000000	0.0000000	0.0000000

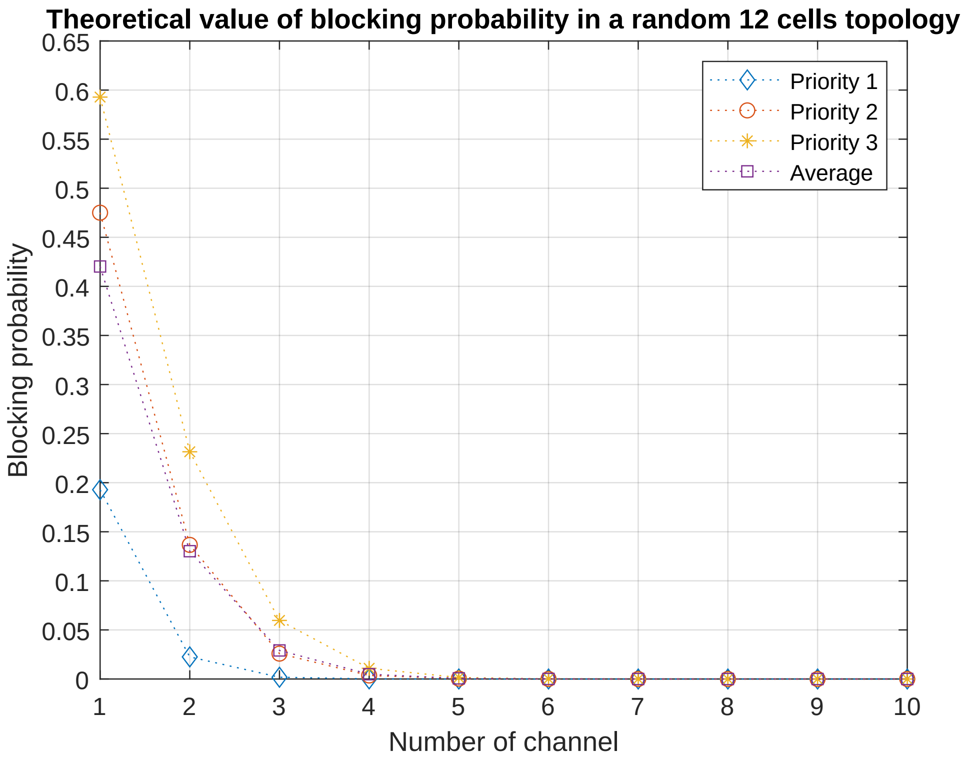


Fig. 4.14 The theoretical calculation of blocking probability of a 12 cells random wireless network

Table 4.8 and Fig. 4.14 show the theoretical calculation results.

Table 4.9 The discrete event simulation result of a 12 cells random cellular network

	Blocking Probability			
Channel number	Priority 1	Priority 2	Priority 3	Average
1	0.1991537	0.4855146	0.5961414	0.3820130
2	0.0255447	0.1603839	0.2493527	0.1153370
3	0.0023370	0.0372628	0.0730744	0.0268720
4	0.0001600	0.0067316	0.0161244	0.0050420
5	0.0000150	0.0010142	0.0028690	0.0008000
6	0.0000000	0.0001080	0.0004398	0.0000980
7	0.0000000	0.0000160	0.0000300	0.0000110
8	0.0000000	0.0000000	0.0000300	0.0000030
9	0.0000000	0.0000000	0.0000000	0.0000000
10	0.0000000	0.0000000	0.0000000	0.0000000

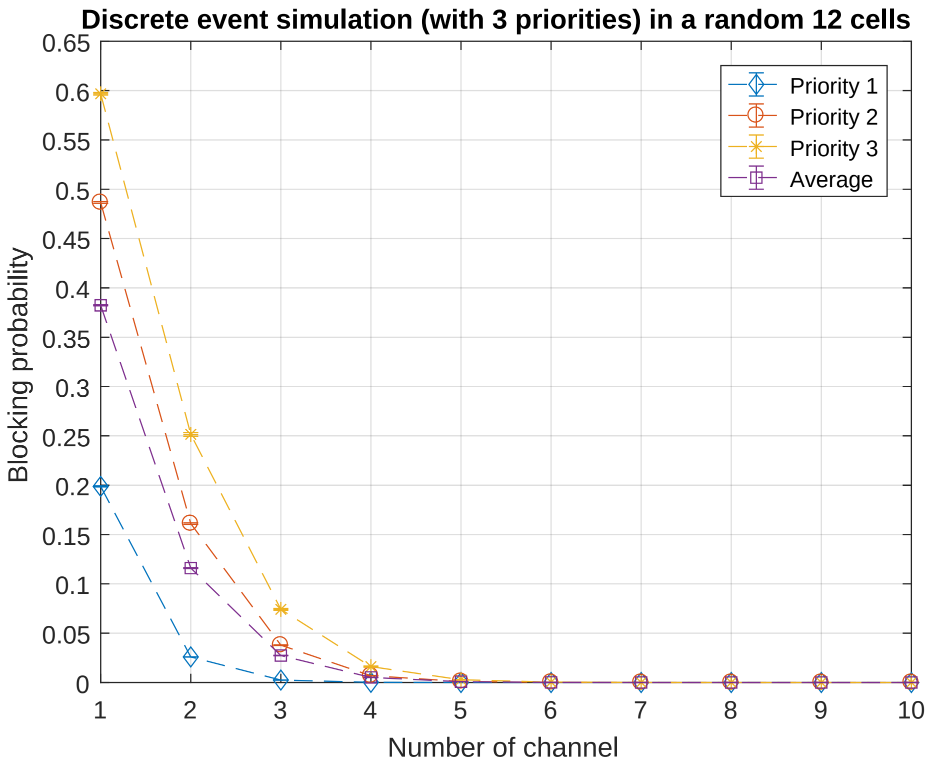


Fig. 4.15 The discrete event simulation blocking probability result of a 12 cells random cellular network

Table 4.9 and Fig. 4.15 show the simulation results. From the simulation results, we see that the blocking probabilities of the three priorities are consistent with the theoretical calculations. Also, the blocking probabilities of each priority in the random cell case are slightly lower than the regular 7 cells case. In the regular 7 cells case, 3 channels are needed to meet the requirement that the highest priority's blocking probability is lower than 5%, while in the random 12 cells case, 2 channels are needed.

4.1.7 ISPs' Cost Saving by Offering Differentiated Services

In this subsection, we investigate the cost-saving that an ISP will achieve when introducing differentiated services. Next, an experiment of changing the proportion of each priority's traffic under different offered load is conducted.

As in the previous experiments/simulations, we keep three priorities. In particular, the QoS requirement (blocking probability) of the first priority is set to be 5%, the second priority's QoS is set to be 10% and the third priority's QoS is set to be 15%. Two extreme values can be calculated, one is the lowest circuit (channels) L needed and the other is the highest circuits (channels) H. The lowest circuit represents the number of circuits needed when only the lowest priority traffic exists, with the blocking probability lower than 15%. On the contrary, the highest circuit is the number of circuits needed when there is only the highest priority traffic, while the blocking probability is less than or equal to 5%. Based on these two extreme values, the lowest number L and highest number H, one can calculate an upper bound of the cost can be saved: $(H-L)/H$.

The aim of this experiment is to show how many circuits (or transmission capacity) can be saved as the proportion of the first priority traffic changes. In this experiment, the proportion of each priority changes as follows: In the beginning, the first priority traffic starts from 1 Erlang, and gradually increases 1 Erlang to the half total offered load, namely, 50% of the total offer load. The second priority traffic is always set to half of the first priority, and the remaining is the third priority traffic. For example, in the case of a total load equal to 100 Erlangs, when the traffic load of the first priority is 20 Erlangs, the traffic load of the second priority goes to 10 Erlangs, and the remaining 70 Erlangs belongs to the third priority.

The flowchart in Fig. 4.16 describes the logic of this experiment:

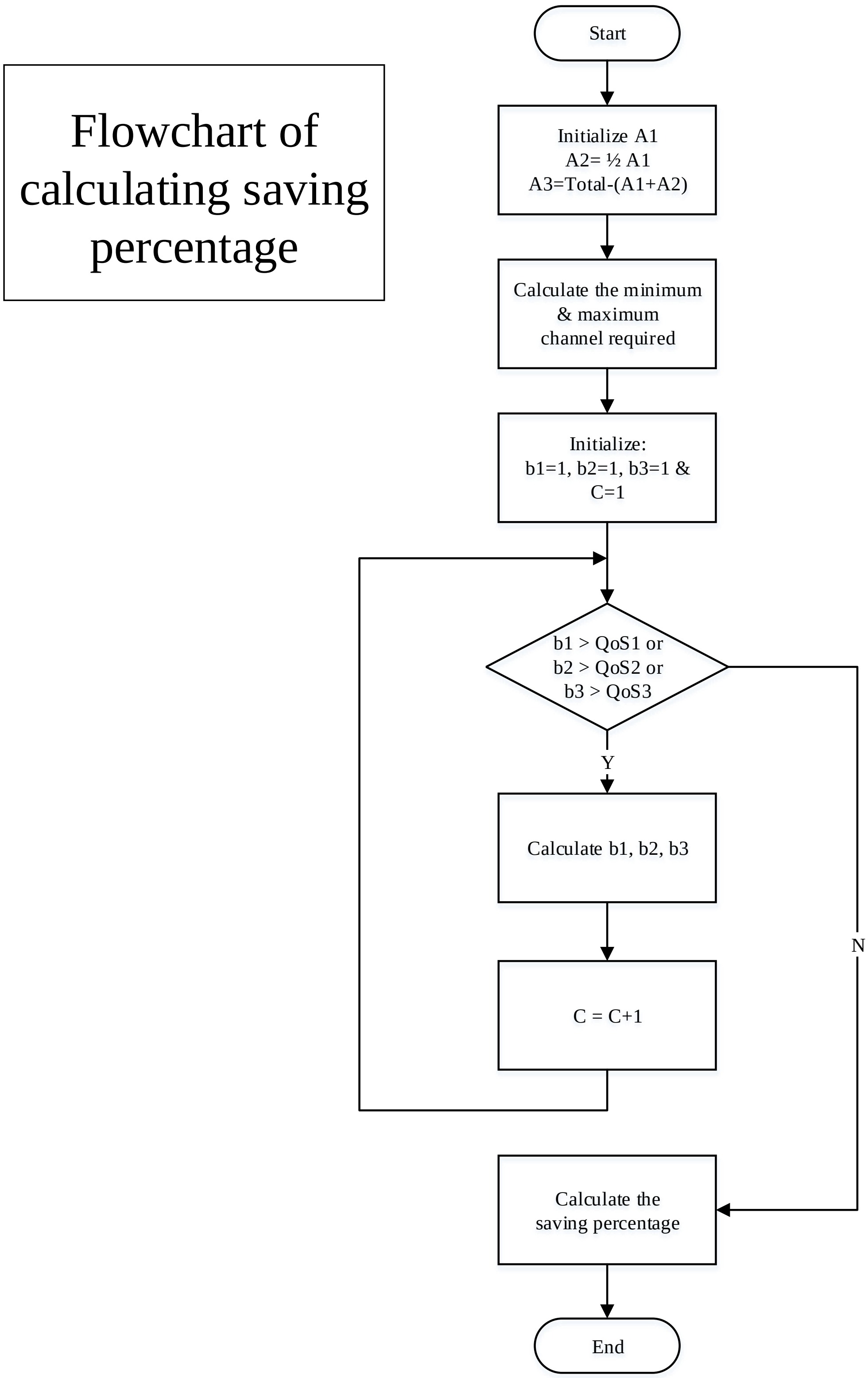


Fig. 4.16 The flowchart of benefit calculation

Experiments are conducted with six different total offer load values: 50, 100, 200, 400, 800, and 1600 Erlangs. The result is shown in Fig. 4.17.

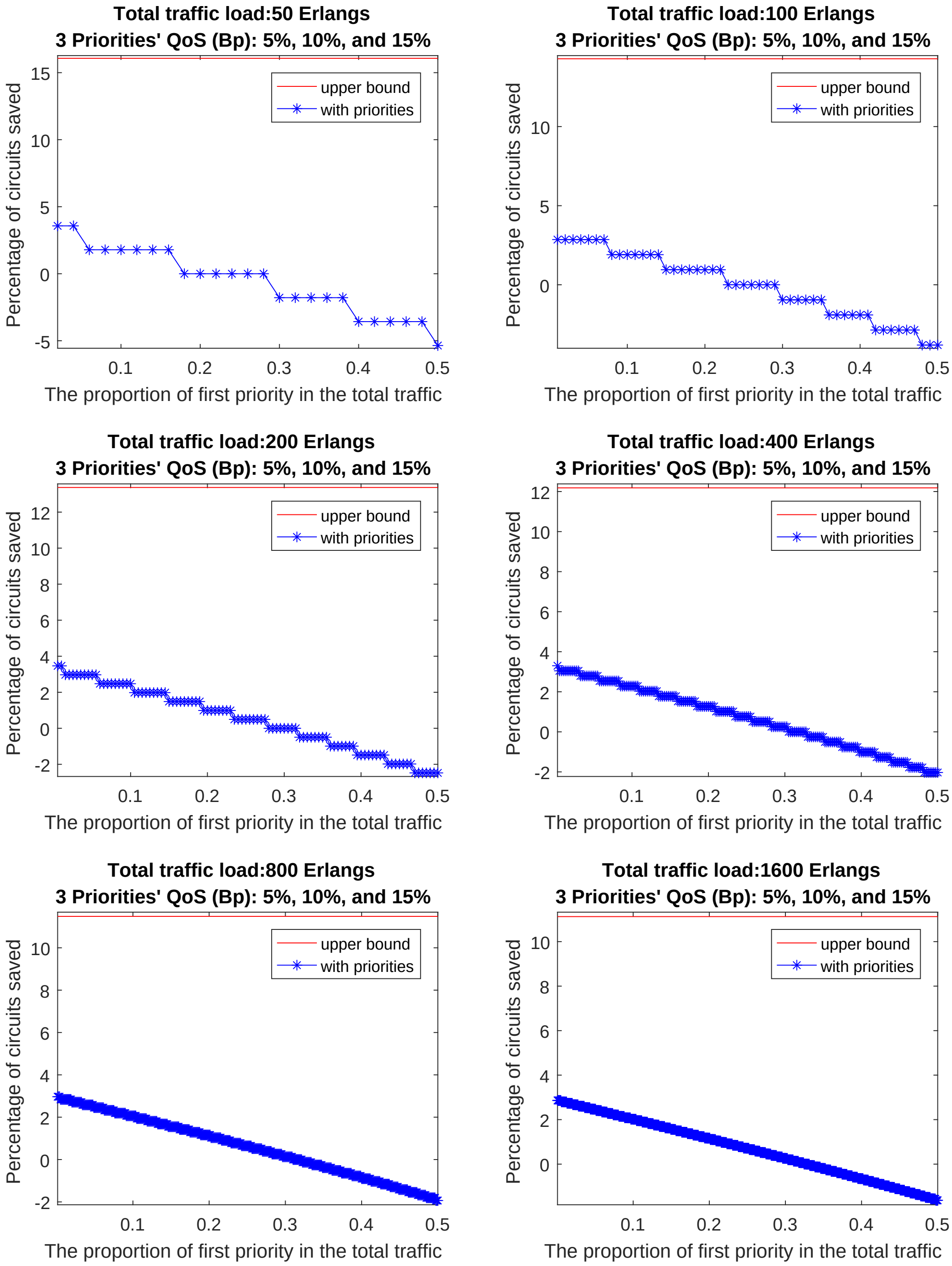


Fig. 4.17 The benefit of circuits (in percentage) saved under different offered load (3 Priority: 5%, 10% and 15%)

As can be seen in Fig. 4.17, when the offered load increases, the percentage of saved circuits go down. This is a consequence of the *Law of large numbers*: When the offered traffic goes up, the uncertainty of the traffic goes down, leading to a decrease in the saved circuits.

The proportion of first priority plays a critical role in this experiment since it decides how many capacities an ISP should deploy. However, the blocking probability, i.e., the QoS for the first priority, is another important factor. Previously, we had set 5% blocking probability as the reference point of the highest priority's QoS, and now we reset the blocking probability as QoS for three priorities: 1% for the highest priority, 5% as the second priority, and 10% as the third priority. We conduct the experiment again, and the result is plotted in Fig. 4.18.

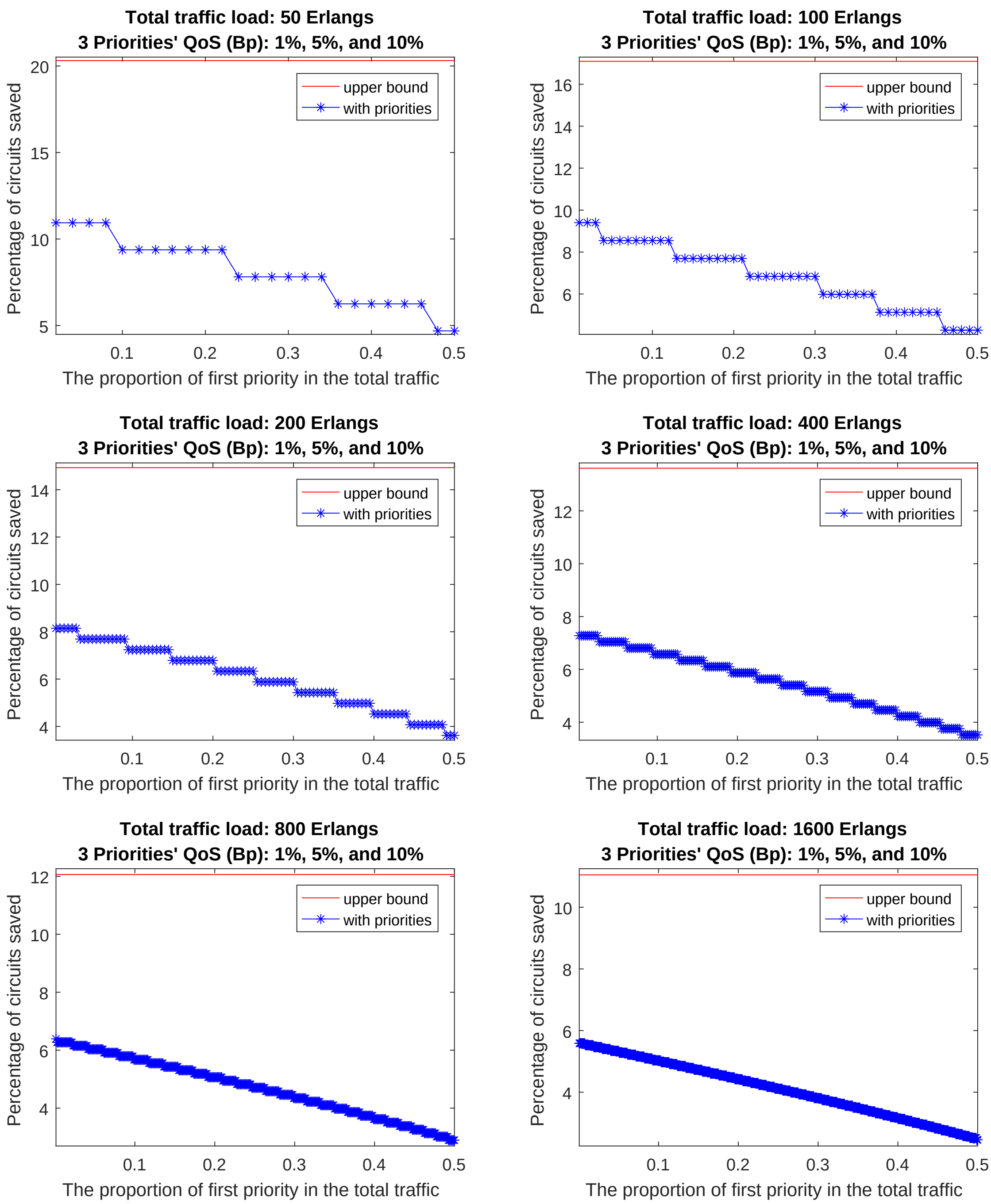


Fig. 4.18 The benefit of circuits (in percentage) saved under different offered load (3 Priority: 1%, 5% and 10%)

From Fig. 4.18, we see that, after increasing the QoS for each class, the percentage of cost-saving increases.

Keeping these QoS for the three priorities: 1%, 5%, and 10%, we further plot the trend of upper-bound with respect to the total offer load. Note that the horizontal axis of Fig. 4.19 is in log-scale. One can see that the benefit of capacity saving decreases when the offered load goes up. For a local ISP whose offered load is 100 Erlangs, theoretically, it can save a maximum of 17% cost by introducing differentiated services, while if a local ISP's offered load is around 1000, then such percentage goes down to 7%.

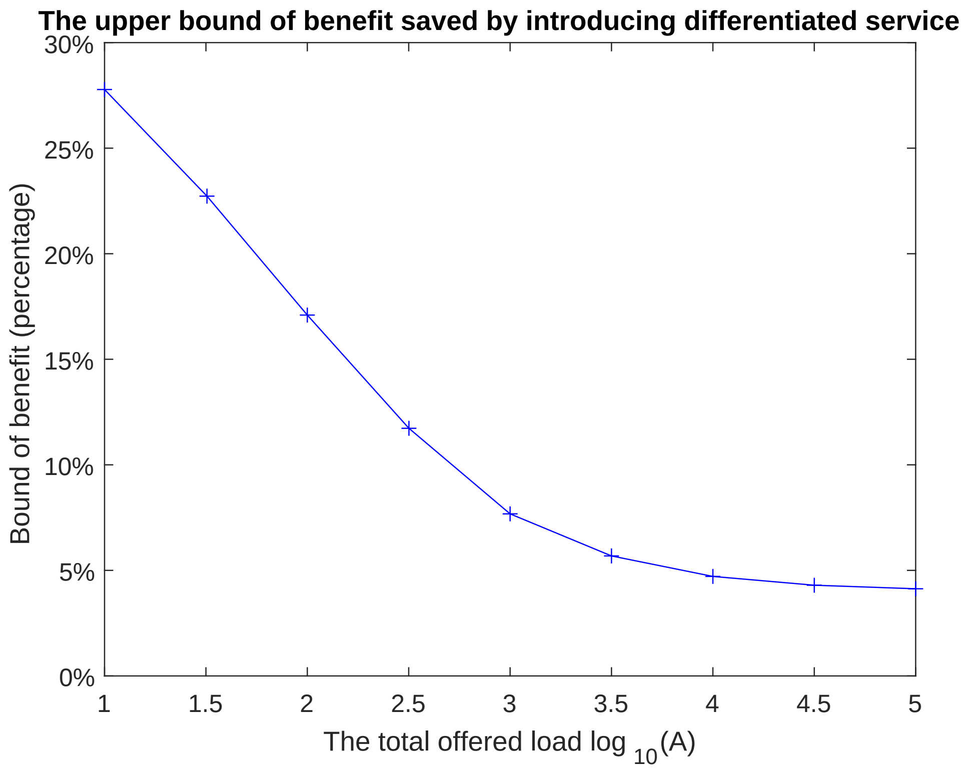


Fig. 4.19 The trend of upper-bound change under different offered load

4.2 Impact of Introducing Differentiated Services on End-users

We have discussed the impact of introducing differentiated services on ISPs. In this section, we investigate the impact of differentiated services from the users' perspective.

4.2.1 Queueing Model and Preliminary

In previous sections, we chose the M/G/*k*/*k* queueing model to study the impact on QoS and cost-saving by introducing differentiated services to the broadband network. Now we change to the M/M/1-PS (processor sharing) model [163, 179–181] to study the impact on the QoE of end-users. The M/M/1-PS model still has the property of being insensitive to the service time distribution. As mentioned previously, the data network does not "block" the newly incoming customers. If the network load is light, each end-user can get the bandwidth it wants, unless the bandwidth exceeds the physical limitation of the network device. For example, the 100Base-T network (Ethernet LAN) cable is able to transmit data at a maximum speed of 100 Mbps. Otherwise, when the network load is high, the capacity is shared by each end-user.

We use the same topology as shown in Fig. 4.1 and study the delay performance.

As discussed in [57, 102], the delay of two classes priority can be described as below:

$$w_1 = \frac{1}{\mu - \lambda_1}, \tag{4.3}$$

and

$$w_2 = \frac{\mu}{\mu - \lambda} w_1 = \frac{\mu}{\mu - \lambda} \frac{1}{\mu - \lambda_1}. \tag{4.4}$$

where λ_1 represents the arrival rate of first class, λ is the total arrival rate, w_1 represents the total delay of first class with highest priority, and w_2 represents the expected waiting time of the second class. Such delay includes both the time waiting in the queue and service time.

4.2.2 Case Study

First, we conduct a simulation of the M/M/1-PS system without priority, the parameters of traffic assumption are the same with previous sections, namely, we set the transmission speed (or capacity) is 10 Gbps, the requests' duration follow the exponential distribution, with the average size of 800 packets, and packet size of 1400 Bytes. The processing rate of this queueing model can be calculated as: $\mu = 10 \times 10^9/(1400 \times 8 \times 800) = 1116.07$

The average delay of the M/M/1-PS system can be theoretically calculated by the following equation:

$$E[D] = \frac{1}{\mu - \lambda} \tag{4.5}$$

From the above equation, we can see that, for the queuing system to be stable, the criteria $\lambda < \mu$, or $\rho = \frac{\lambda}{\mu} < 1$ must be met, otherwise the formula is not valid. In the simulation, λ changed from 100 to 1000 Erlang with a step size of 100 Erlang. The number of total arrivals is set to 10^6 (1 million) requests, and the simulation is run 6 times to get a 95% confidence interval.

ρ	**Theoretical value**	**Simulation result**
0.0896	0.0009842	0.0009860
0.1792	0.0010916	0.0010964
0.2688	0.0012254	0.0012265
0.3584	0.0013965	0.0014023
0.448	0.0016232	0.0016172
0.5376	0.0019377	0.0019292
0.6272	0.0024034	0.0024082
0.7168	0.0031638	0.0031639
0.8064	0.0046281	0.0046423
0.896	0.0086154	0.0085819

Table 4.10 The delay time (second) of M/M/1-PS system under different offered load (ρ)

Fig. 4.20 shows the results of both analytical calculation and simulation.

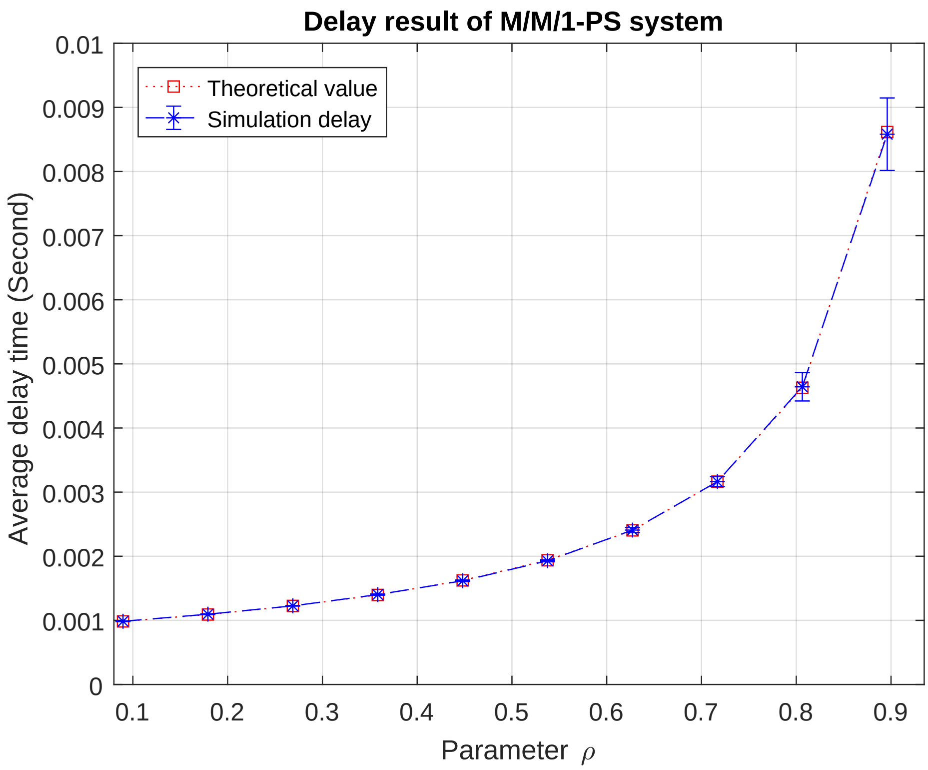


Fig. 4.20 The analytical and simulation results of M/M/1-PS system

Next, we conduct the simulation of the M/M/1-PS system with 3 priorities. We set 40% of the total traffic as priority 1, 50% of the total traffic as priority 2, and the remaining 10% as priority 3.

Since the highest priority is the most sensitive to delay, Table 4.11 and Fig. 4.21 give the theoretical value and experimental results of priority 1. The average delay (the delay without priority) is also provided as a comparison.

ρ	**Theoretical value (without Priority)**	**Theoretical value (Priority 1)**	**Simulation result (Priority 1)**	**Saving on latency (in percentage)**
0.0896	0.0009842	0.00092931	0.00094652	3.88%
0.1792	0.0010916	0.00096518	0.0010011	7.77%
0.2688	0.0012254	0.0010039	0.0010631	13.42%
0.3584	0.0013965	0.0010459	0.001135	19.28%
0.448	0.0016232	0.0010916	0.0012131	25.73%
0.5376	0.0019377	0.0011415	0.0013012	32.84%
0.6272	0.0024034	0.0011961	0.0014119	41.77%
0.7168	0.0031638	0.0012562	0.0015239	51.26%
0.8064	0.0046281	0.0013226	0.001669	64.34%
0.896	0.0086154	0.0013965	0.0018509	78.63%

Table 4.11 The delay time (second) of MM1-PS system with priority under different offered load (ρ)

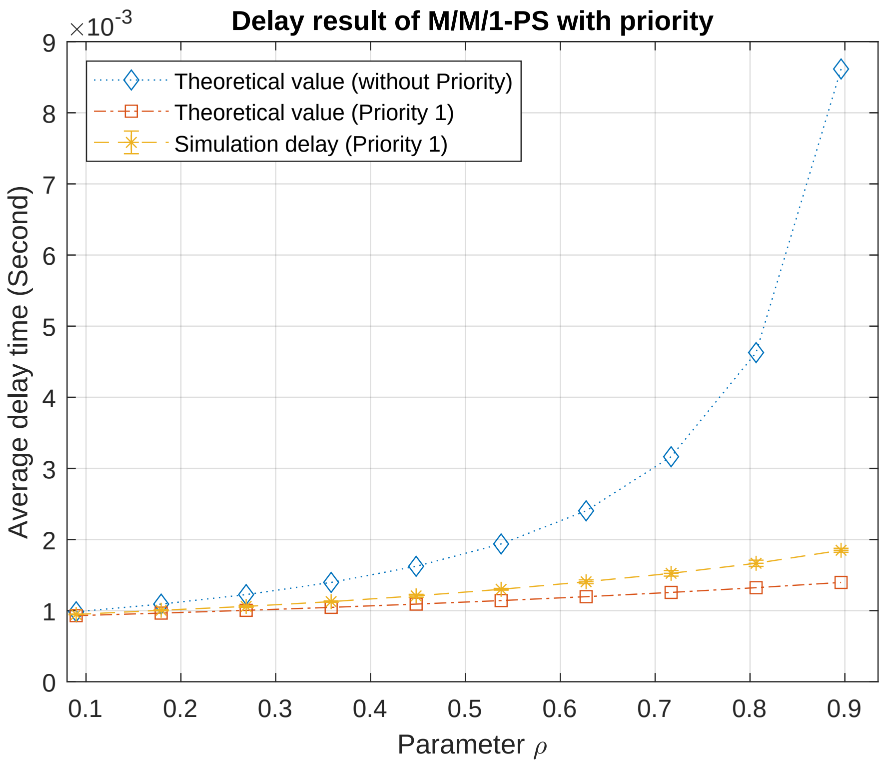


Fig. 4.21 The simulation results of MM1PS system with priorities

We see from the results that when the traffic load is small ($\rho \approx 0.1$), the highest priority can save 3.88 % of the waiting time, and when the traffic load is high ($\rho \approx 0.9$), the highest priority can save nearly 80% of the waiting time. This means that when the traffic is close to the critical load (when $\rho \approx 1$), high priority ensures that the performance is not significantly degraded, which is particularly important for latency-sensitive services.

4.3 Comparative Analysis of Service Differentiation Based on Real Datasets

In the previous sections of this chapter, we investigated the impact of differentiated services on both ISPs and end-users, based on the assumption that Poisson arrivals and the duration are exponentially distributed (or Pareto distributed). In this section, we investigate the impact of differentiated services on network services from the dimension of service types (protocols). More importantly, in order to enhance persuasion and practicality, the following analysis will be based on real network datasets.

The basic idea of implementing differentiated services based on service protocols is that different service scenarios have different transmission requirements. As described in Chapter 3, TCP is a connection-oriented protocol that provides error checking and retransmission mechanisms. TCP has a flow control mechanism for each flow, which can adjust the sending rate according to the receiver's situation. UDP is a connectionless protocol with no error checking mechanism and flow control. It is usually used for the transmission of certain time-sensitive data, such as video streaming, real-time broadcasting, and VoIP. It should be pointed out that using various protocols to transmit data in networks has been accepted by the public for a long time, and has not been challenged by net neutrality proponents. Such practice of transmitting data using different transmission protocols, such as TCP and UDP, can be regarded as a social norm. Therefore, it may be acceptable to prioritize services based on different service protocols. UDP is usually used for time-sensitive services which can be given higher priority. Services supported by TCP do not require hard real-time performance (such as HTTP browsing and e-mail transmission) can be given lower priority. It is worth noting that in addition to the transmission protocols TCP and UDP, there is a higher-level control protocol ICMP (Internet Control Message Protocol) in the network, which is responsible for sending error messages and operation instructions to indicate the success and failure of communicating with another IP address. Such ICMP should be assigned the highest priority in the network.

4.3.1 Datasets

For research purpose, there are numerous datasets that can be download from the website and which captures the real traffic through certain ISPs.

In order to be widely representative of time and region, we use four groups of datasets from two websites: "http://mawi.wide.ad.jp/mawi/" and "https://data.caida.org/datasets".

For the first website, we chose the datasets from its sample-point F, where data packets that are transmitted from the WIDE project to the upstream ISP are recorded daily. "WIDE" is a Japan-based Internet technologies research project started in 1988.

We chose the following five datasets for group 1:

Group 1:

1. "WIDE_20150917_130700"
2. "WIDE_20180621_130000"
3. "WIDE_20200315_130140"
4. "WIDE_20200316_130640"
5. "WIDE_20200316_131000"

The first and second datasets were collected on September 17, 2015, and on June 21, 2018, respectively. The last three datasets were collected on March 15th and 16th, 2020.

For the second website CAIDA (Center for Applied Internet Data Analysis), we chose the following three groups of datasets:

Group 2: https://data.caida.org/datasets/passive-2015/equinix-chicago.dirA/

1. "20150219-131200/UTC.anon.pcap"
2. "20150521-130800/UTC.anon.pcap"
3. "20150917-130800/UTC.anon.pcap"
4. "20151217-131800/UTC.anon.pcap"

Group 3: https://data.caida.org/datasets/passive-2016/equinix-chicago.dirA/

1. “20160121-132200/UTC.anon.pcap”,
2. “20160218-131200/UTC.anon.pcap”,
3. “20160317-131000/UTC.anon.pcap”,
4. “20160406-131000/UTC.anon.pcap”,

Group 4: https://data.caida.org/datasets/passive-2018/equinix-nyc.dirA/

1. “20180315-130200/UTC.anon.pcap”,
2. “20180419-131300/UTC.anon.pcap”,
3. “20180517-133800/UTC.anon.pcap”,
4. “20180621-134500/UTC.anon.pcap”,
5. “20180719-132300/UTC.anon.pcap”,
6. “20180816-133200/UTC.anon.pcap”

The second group of datasets was collected in the city of Chicago on February 19, May 21, September 17, and December 17, 2015. The third group of datasets was also collected in the city of Chicago on January 21, February 18, March 17, and April 6, 2016. The fourth group of datasets was collected in New York City at the time of March 15, April 19, May 17, June 21, July 19, and August 16, 2018.

In order to protect privacy, the payload of all data packets has been removed, the original IP address, the target IP address, and the transmission protocol used were retained.

4.3.2 Datasets Analysis Tool

We use a network analysis tool named Wireshark [182], which is an open-source packet-level analyzer.

The transmission protocols are set as filters, all the data packets were extracted and classified into three groups: ICMP, UDP, and TCP. The time granularity of the analysis is set to milliseconds, and the unit of data is bits.

We also plot the graph of each dataset to give a visual impression of each dataset. Please refer to appendix A.

4.3.3 Three Scenarios for Comparative Study

To be consistent with previous sections, we still use the hard blocking probability as an indicator to examine the impact on the QoS.

Video streaming is generally considered to have the highest QoS requirements. It is believed that an acceptable blocking probability of video stream in an IP network is below 0.25% [183]. In the following analysis, we use 0.25% as a benchmark to dimension the transmission capacity of the ISP. The comparative analysis is carried on three scenarios:

1. The NN scenario
2. The dedicated line scenario
3. The preemptive priority scenario

In the NN scenario, all traffic goes through the ISP in a first-come-first-served (FCFS) manner. If the ISP's bandwidth is insufficient to transmit all traffic, the overflow traffic will be discarded.

The dedicated line scenario is similar to using VPN technology. The original intention of VPN technology is for information encryption. The side effect of a VPN is the improvement of network service quality. Some VPNs are implemented by renting dedicated lines, unlike public networks that serve a large number of end-users. Due to the limited number of end-users, VPNs improve network service quality, such as more stable connections and shorter delays.

The preemptive priority scenario is similar to Section 4.1, we also set three priorities: the first (highest) priority is assigned to ICMP packets, the second priority is assigned to UDP packets, and the third priority is assigned to TCP packets.

4.3.4 Capacity Dimensioning

The capacity needed for each data set is calculated by the follow equation:

$$Capacity = mean(Traffic\ Volume) + 3 \times std(Traffic\ Volume) \tag{4.6}$$

where $mean()$ is the average value, and $std()$ represents the standard deviation.

It is well known that Gaussian random variables obey the so-called 68-95-99.7% rule, which can be expressed as follows:

$$Pr(\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 0.6827$$
$$Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$$
$$Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$$

where X is the Gaussian $N(\mu, \sigma^2)$ variable.

The rule explains that for any random variable that follows a normal distribution, the probability that it falls within the region $[\mu - 1\sigma, \mu + 1\sigma]$ is 68.27%, etc.

Inspired by this empirical rule, we try to dimension the capacity of ISP based on the mean and standard deviation of the traffic volume of each dataset. However, we exam all data sets with the Lillietest [184] in MATLAB [185] and found that none of the datasets follows a normal distribution.

4.3.5 Comparative Analysis in Three Scenarios

Although the traffic volume of datasets does not follow the normal distribution, we still follow the method of using the average value of traffic plus several times of the standard deviation to dimension the capacity.

The comparative study is carried on three scenarios:

1. In the First "NN FCFS" scenario, we use the average value of the traffic volume plus 1 standard deviation and then derive the blocking probability of three classes of traffic under this capacity. If this capacity is not sufficient to meet the blocking probability of any protocols less than or equal to 0.25%, we increase the value of the standard deviation in steps of 0.1 until the blocking probability of all protocol traffic is less than 0.25%. The algorithm is shown in Algorithm 1.

2. In the second "Dedicated line" scenario, we keep the same total capacity C calculated in the previous step, and used the same algorithm to calculate the dedicated capacity C_U and relevant k_U value for UDP traffic, so that the blocking probability Pb_U is less or equal to 0.25%. The remaining capacity $C_R = C - C_U$ is then used to serve the ICMP and TCP traffic, and we can derive the blocking probability of ICMP and TCP, respectively.

3. In the third "Preemptive priority" scenario, we still use the same capacity C as the total transmission capacity. Within each millisecond time unit, the transmission capacity will be preferentially allocated to ICMP, then to UDP, and finally to TCP. We will get the blocking probability of each of these three groups.

Algorithm 1: Find the specific capacity C for each dataset

Result: Find capacity C and relevant k value

$k = 0.9$;

Blocking probability $P_b = 1$;

while P_b *is greater than 0.25%* **do**

  increase k by 0.1;

  C = mean(Traffic Volume) + $k \times$ std(Traffic Volume);

  calculate the blocking probability under C;

end

Combining the above three scenarios, we can make a comparative analysis. The results are shown in Tables 4.12, 4.13, and 4.14, respectively.

Dataset	Net Neutrality Scheme (FIFO)		
	k value	Total Capacity (bps)	BP (all)
WIDE_20200315_130140	4.6	1135694447	0.2260%
WIDE_20200316_130640	2.7	1556114015	0.2210%
WIDE_20200316_131000	2.9	1618127364	0.2420%
WIDE_20180621_130000	3.3	9033657191	0.0833%
WIDE_20150917_132600	3.3	4567636161	0.2350%
Chicago_20150219_131200	3	5038689436	0.2300%
Chicago_20150521-130800	3.1	4968595418	0.2117%
Chicago_20150917-130800	3.2	4585662701	0.2467%
Chicago_20151217-131800	3.4	5426973895	0.2417%
Chicago_20160121-132200	3.7	6618986532	0.2367%
Chicago_20160218-131200	5	8598858480	0.2397%
Chicago_20160317-131000	4.3	5817656895	0.2383%
Chicago_20160406-131000	3.8	6314069935	0.2476%
Nyc_20180315-130200	3.1	7551260026	0.2317%
Nyc_20180419-131300	3.6	8748458664	0.2450%
Nyc_20180517-133800	2.6	9103611858	0.0000%
Nyc_20180621-134500	3	9073991776	0.0000%
Nyc_20180719-132300	3.2	8785375551	0.2450%
Nyc_20180816-133200	3	8739267603	0.2167%

Table 4.12 The NN scheme discards overflow packets

Dataset	**Dedicated Line Scheme**				
	k_U value	Total Capacity (bps)	UDP Capacity (bps)	BP (ICMP + TCP)	BP (UDP)
WIDE_20200315_130140	6.6	1135694447	77677413.93	0.6940%	0.2500%
WIDE_20200316_130640	7	1556114015	343652361.4	0.0040%	0.2470%
WIDE_20200316_131000	3.8	1618127364	443909614.9	1.8640%	0.2500%
WIDE_20180621_130000	3.2	9033657191	988293149	0.5300%	0.2467%
WIDE_20150917_132600	3.2	4567636161	276026492.4	0.3300%	0.2333%
Chicago_20150219_131200	4.3	5038689436	555596094.9	0.4983%	0.2383%
Chicago_20150521-130800	3.5	4968595418	407819141.9	0.3683%	0.2500%
Chicago_20150917-130800	3.3	4585662701	296208120.3	0.3783%	0.2450%
Chicago_20151217-131800	3.1	5426973895	461568267	0.3683%	0.2233%
Chicago_20160121-132200	2.8	6618986532	414389877.9	0.2800%	0.2400%
Chicago_20160218-131200	3.5	8598858480	421309917.5	0.2737%	0.2193%
Chicago_20160317-131000	5	5817656895	616813883.7	0.3383%	0.2500%
Chicago_20160406-131000	3	6314069935	355585945.6	0.2807%	0.2284%
Nyc_20180315-130200	3.3	7551260026	1013095406	0.5250%	0.2467%
Nyc_20180419-131300	2.8	8748458664	835613638	0.3900%	0.2433%
Nyc_20180517-133800	3.7	9103611858	1092725277	1.0000%	0.2467%
Nyc_20180621-134500	3	9073991776	963731056.2	0.5217%	0.2350%
Nyc_20180719-132300	2.8	8785375551	1020621984	0.4683%	0.2217%
Nyc_20180816-133200	2.7	8739267603	1153700382	0.4567%	0.2317%

Table 4.13 The Dedicated line scheme discards overflow packets

Dataset	Preemptive Priority Scheme			
	Total Capacity (bps)	BP (ICMP)	BP (UDP)	BP (TCP)
WIDE_20200315_130140	1135694447	0.0000%	0.0000%	0.2710%
WIDE_20200316_130640	1556114015	0.0000%	0.0000%	0.2690%
WIDE_20200316_131000	1618127364	0.0000%	0.0000%	0.3580%
WIDE_20180621_130000	9033657191	0.0000%	0.0000%	0.0883%
WIDE_20150917_132600	4567636161	0.0000%	0.0000%	0.2867%
Chicago_20150219_131200	5038689436	0.0000%	0.0000%	0.2450%
Chicago_20150521-130800	4968595418	0.0000%	0.0000%	0.2383%
Chicago_20150917-130800	4585662701	0.0000%	0.0000%	0.2750%
Chicago_20151217-131800	5426973895	0.0000%	0.0000%	0.3117%
Chicago_20160121-132200	6618986532	0.0000%	0.0000%	0.2883%
Chicago_20160218-131200	8598858480	0.0000%	0.0000%	0.2822%
Chicago_20160317-131000	5817656895	0.0000%	0.0000%	0.3067%
Chicago_20160406-131000	6314069935	0.0000%	0.0000%	0.2911%
Nyc_20180315-130200	7551260026	0.0000%	0.0000%	0.2800%
Nyc_20180419-131300	8748458664	0.0000%	0.0000%	0.2617%
Nyc_20180517-133800	9103611858	0.0000%	0.0000%	0.0000%
Nyc_20180621-134500	9073991776	0.0000%	0.0000%	0.0000%
Nyc_20180719-132300	8785375551	0.0000%	0.0000%	0.2550%
Nyc_20180816-133200	8739267603	0.0000%	0.0000%	0.2350%

Table 4.14 The Preemptive priority scheme discards overflow packets

In reality, TCP retransmits unsuccessful/lost packets, the following simulation considers the TCP retransmission mechanism and recalculated the capacity. The results are listed in Tables 4.15 - 4.17.

Dataset	Net Neutrality Scheme (FIFO)		
	k value	Total Capacity (bps)	BP (all)
WIDE_20200315_130140	4.7	1155297024	0.2040%
WIDE_20200316_130640	2.8	1593980682	0.1920%
WIDE_20200316_131000	3.1	1693757514	0.2170%
WIDE_20180621_130000	3.3	9033657191	0.0883%
WIDE_20150917_132600	3.4	4648647295	0.2433%
Chicago_20150219_131200	3	5038689436	0.2450%
Chicago_20150521-130800	3.1	4968595418	0.2383%
Chicago_20150917-130800	3.3	4668473735	0.2283%
Chicago_20151217-131800	3.7	5671229225	0.2383%
Chicago_20160121-132200	3.8	6730067601	0.2500%
Chicago_20160218-131200	5.2	8808409568	0.2499%
Chicago_20160317-131000	4.7	6137692309	0.2383%
Chicago_20160406-131000	4.1	6608803522	0.2354%
Nyc_20180315-130200	3.2	7676368898	0.2317%
Nyc_20180419-131300	3.7	8879166503	0.2150%
Nyc_20180517-133800	2.6	9103611858	0.0000%
Nyc_20180621-134500	3	9073991776	0.0000%
Nyc_20180719-132300	3.3	8918341158	0.2100%
Nyc_20180816-133200	3	8739267603	0.2350%

Table 4.15 The NN scheme retransmits overflow TCP packets

Dataset	**Dedicated Line Scheme**				
	k_U value	Total Capacity (bps)	UDP Capacity (bps)	BP (ICMP + TCP)	BP (UDP)
WIDE_20200315_130140	6.6	1155297024	77677414	0.7200%	0.2500%
WIDE_20200316_130640	7	1593980682	343652361	1.6860%	0.2470%
WIDE_20200316_131000	3.8	1693757514	443909615	1.8930%	0.2500%
WIDE_20180621_130000	3.2	9033657191	988293149	0.5750%	0.2467%
WIDE_20150917_132600	3.2	4648647295	276026492	0.3350%	0.2333%
Chicago_20150219_131200	4.3	5038689436	555596095	0.5433%	0.2383%
Chicago_20150521-130800	3.5	4968595418	407819142	0.4283%	0.2500%
Chicago_20150917-130800	3.3	4668473735	296208120	0.3483%	0.2450%
Chicago_20151217-131800	3.1	5671229225	461568267	0.3367%	0.2233%
Chicago_20160121-132200	2.8	6730067601	414389878	0.3183%	0.2400%
Chicago_20160218-131200	3.5	8808409568	421309918	0.2924%	0.2193%
Chicago_20160317-131000	5	6137692309	616813884	0.3383%	0.2500%
Chicago_20160406-131000	3	6608803522	355585946	0.2772%	0.2284%
Nyc_20180315-130200	3.3	7676368898	1013095406	0.5367%	0.2467%
Nyc_20180419-131300	2.8	8879166503	835613638	0.3750%	0.2433%
Nyc_20180517-133800	3.7	9103611858	1092725277	1.1083%	0.2467%
Nyc_20180621-134500	3	9073991776	963731056	0.5700%	0.2350%
Nyc_20180719-132300	2.8	8918341158	1020621984	0.4417%	0.2217%
Nyc_20180816-133200	2.7	8739267603	1153700382	0.5117%	0.2317%

Table 4.16 The Dedicated line scheme retransmits overflow TCP packets

Dataset	**Preemptive Priority Scheme**			
	Total Capacity (bps)	BP (Control)	BP (UDP)	BP (TCP)
WIDE_20200315_130140	1155297024	0.0000%	0.0000%	0.2040%
WIDE_20200316_130640	1593980682	0.0000%	0.0000%	0.1920%
WIDE_20200316_131000	1693757514	0.0000%	0.0000%	0.2170%
WIDE_20180621_130000	9033657191	0.0000%	0.0000%	0.0883%
WIDE_20150917_132600	4648647295	0.0000%	0.0000%	0.2433%
Chicago_20150219_131200	5038689436	0.0000%	0.0000%	0.2450%
Chicago_20150521-130800	4968595418	0.0000%	0.0000%	0.2383%
Chicago_20150917-130800	4668473735	0.0000%	0.0000%	0.2283%
Chicago_20151217-131800	5671229225	0.0000%	0.0000%	0.2383%
Chicago_20160121-132200	6730067601	0.0000%	0.0000%	0.2500%
Chicago_20160218-131200	8808409568	0.0000%	0.0000%	0.2499%
Chicago_20160317-131000	6137692309	0.0000%	0.0000%	0.2383%
Chicago_20160406-131000	6608803522	0.0000%	0.0000%	0.2354%
Nyc_20180315-130200	7676368898	0.0000%	0.0000%	0.2317%
Nyc_20180419-131300	8879166503	0.0000%	0.0000%	0.2150%
Nyc_20180517-133800	9103611858	0.0000%	0.0000%	0.0000%
Nyc_20180621-134500	9073991776	0.0000%	0.0000%	0.0000%
Nyc_20180719-132300	8918341158	0.0000%	0.0000%	0.2100%
Nyc_20180816-133200	8739267603	0.0000%	0.0000%	0.2350%

Table 4.17 The Preemptive priority scheme retransmits overflow TCP packets

From these results, whether the overflow packets are discarded or not, we find that in the dedicated line scheme, the UDP traffic gets the dedicated capacity for its transmission, so that the capacity of ICMP and TCP is squeezed, hence the performance of ICMP and TCP is degraded. While keeping the total capacity the same, the preemptive scheme shows that the blocking probabilities of the ICMP and the UDP are controlled to be zero, and the blocking probability of TCP is still acceptable. This shows that service differentiation based on service protocol is feasible, without causing discomfort to end-users.

4.4 Summary

This chapter studies the impact of introducing differentiated services into data networks from two different perspectives, namely user groups (paid prioritization) and data type (transmission protocol). We can consider four options:

1. The first is a neutral network, which does not prioritize user groups or data types.
2. The second is to give different priorities to different user groups based on their willingness to pay: higher-paying users get higher priority, which is consistent with popular business concepts.
3. Third, different priorities are given to different types of data traffic.
4. The fourth is based on the priority of payment and gives different priorities to different data types at the same time. That is, even in user groups with the same priority, priorities are assigned based on real-time requirements of different data types.

For a clear view, we use the Table 4.18 to illustrate the four options:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Based on user group (paid prioritization)</td></tr>
<tr><td>NO</td><td>YES</td></tr>
<tr><td rowspan="2">Base on service type</td><td>NO</td><td>Option 1: Neutral Net</td><td>Option 2: User group-prioritized Net</td></tr>
<tr><td>YES</td><td>Option 3: Service-prioritized Net</td><td>Option 4: User and service-prioritized Net</td></tr>
</table>

Table 4.18 Four combinations of differentiated services

Option 1 describes most existing data networks, where all data is treated “equally”, and is the way that network neutrality proponents want.

Option 2 is a paid priority scheme, and much literature has concluded that it increases ISP’s investment incentives. Of course, it is also the source of controversy about network neutrality. Our simulation results show the impact of differentiated services: For ISPs, differentiated services not only provide different QoE for different users, but also bring cost-saving advantages. For end-users, high priority ensures that performance will not degrade significantly under high network load.

Inspired by the discussion of DiffServ in the previous chapter, we proposed Option 3, which gives different priorities to different types of traffic. We tested this option based on real data sets. The results show that, with the same transmission capacity as Option 1, the blocking probability of services with real-time performance decreases, so the end-users’ experience is improved. Such differentiation of the data packets by transmission protocols can be implemented by setting parameters at the egress/ingress router side of the ISP, which will not incur too much operation cost to the ISPs. On the other hand, it improves the users’ experience, Therefore, this option may be a viable method for resolving network neutrality disputes.

For option 4, it is foreseeable that if the same user group is further allowed to be differentiated on the basis of option 3, ISPs will have more investment incentives to invest in their network facilities, which will improve social welfare. This may be an option for ISPs to provide differentiated services in the future.

In summary, this chapter makes a quantitative analysis of differentiated services from the perspective of network performance. In addition to the theoretical feasibility analysis, listening to the opinions of stakeholders is also decisive for the establishment of public policy. In the next chapter, we present a social survey to solicit opinions on net neutrality.

Chapter 5

A Survey Experiment of Net Neutrality Policy

The Hong Kong Special Administrative Region of China is known for its open trade and high freedom of speech, with the Internet playing a critical role in both trade and expression. There is no doubt that Internet connection speeds have a profound impact on maintaining Hong Kong's competitiveness.

According to recent statistics [186], Hong Kong's average download speed in 2018 was 26.45 Mbps, ranking 4th in the Asia Pacific region and 19th in the world. However, this average speed cannot describe the full picture. The coverage of the network infrastructure is uneven, resulting in insufficient network services in some areas of Hong Kong. Outlying islands far from the downtown area, including Lantau Island and Lamma Island, suffer very low access rates [187]. Network management policy plays a decisive role in how to improve the investment incentives of network operators, while taking into account the fairness of market competition.

Unlike some countries in Europe and Latin America [7, 41, 44], Hong Kong has not yet legislated on net neutrality at the time of writing. Listening to public opinion is very necessary to make a proper public policy. In November 2018, we conducted a policy survey experiment on “smart cities” in Hong Kong. Autonomous vehicles and net neutrality were the two main themes of this survey. In this chapter, we will introduce the net neutrality part of this survey and analyze the polling data. For the full content of the questionnaire, see Appendix B.

Net neutrality has been under debate for decades in western countries, and several public surveys on network management policy have been conducted, such as in the United States [188–190] and Europe [191]. This work, as far as we know, is the first public survey on network management policy in Hong Kong. Using survey experiments in network management policy polling is rare. A similar work can be found in [192], which is a survey experiment on consumers’ video-on-demand market in the United States.

5.1 Research Design

Experiment research has been used to analyzed policy issues, and has more recently been adopted in public administration research where it has proved to be effective and fruitful [193, 194]. Survey experiments, or, experimental vignette method (EVM) is one of the main experimental methodologies adopted. The basic principle of EVM is to isolate the causal relationships by randomly assign respondents to statistically equivalent groups, thereby avoiding potential problems (e.g., variable bias), which often leads to deterioration on multivariate regression.

EVM has been proved to be particularly useful when exercising control of independent variables is needed [195]. In this survey experiment, we focus on dimensions of performance, i.e., equity and efficiency.

5.1.1 Research Setting

In the design of EVM, we carefully design realistic scenarios to evaluate dependent variables including equity and efficiency of the broadband market in Hong Kong. The equity variable concerns the fairness of competition in the broadband market, where “zero-rating” is used as the entry point. The efficiency variable examines the quality of service in the broadband market through “bandwidth throttling”.

5.1.2 Participants and Procedures

Participants:

The survey was conducted through “My Citizens Panel”, which is hosted by the Laboratory for Public Management and Policy at City University of Hong Kong. The survey was opened in November 2018. Participants were contacted by email, and those wishing to partake in the study responded to the email to access the survey which was designed in Qualtrics. At the time of writing, 497 effective respondents have submitted the questionnaire (See Fig. 5.1). The survey was conducted anonymously, yet respondents needed to provide some basic socio-demographic information. There were 429 respondents in the age range from 20 to 49, which is over 86% of the total number, and 299 respondents are full-time employed, which is 60% of the total.

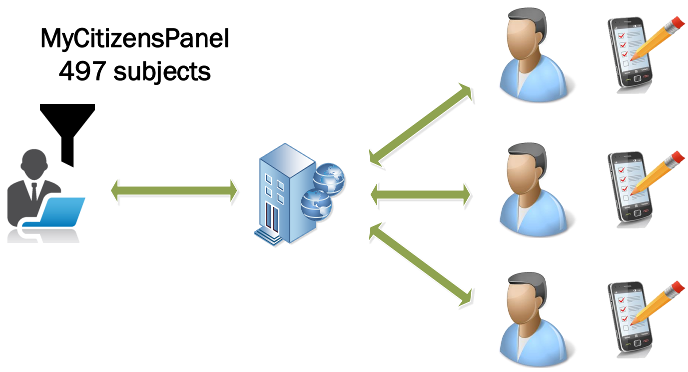


Fig. 5.1 The polling procedure

Procedures:

Considering the complexity of the concept of network neutrality, we gave a series of conceptual explanations about network usage and operation at the very beginning of the questionnaire. These include “network operators”, “content providers”, “zero-rating”, “network congestion”, “bandwidth throttling”, “investment incentives”, “network resource allocation”, “justification for service prioritization”, and “net neutrality”.

After the explanations of terminologies, we prepared preliminary questions to prevent priming effects. These questions investigated the network type (wireline or wireless network), the expenditure per month of using the network, the average time usage level, and satisfaction with one’s ISP.

Given the complexity and length of the survey, the respondents were compensated HKD $50 (US$6.4) for their participation.

5.1.3 Experiment Vignettes

Following the ideas of designing vignettes [193, 196], we set “zero-rating” as equity and “bandwidth throttling” as efficiency. As discussed in Chapter 2, some big content providers tend to cooperate with ISPs to subsidize their subscribers, even they claim that they are proponents of net neutrality. Such “zero-rating” practice obviously violates the fair competition rule. Therefore “zero-rating” is adopted here to test respondents’ sensitivity of fairness. Also, as analyzed in Chapter 4, we conclude that applying prioritized service in broadband network not only increase the investment incentives of ISPs but also optimize end-users’ quality of experience. To explain the prioritized service in the broadband network, we use the term “Bandwidth throttling under and not under network congestion” in the terminologies and related information part.

After the respondents finished the preliminary questions, they were randomly assigned a vignette to read. Specifically, we set that vignette 1 allows neither “zero-rating” nor "bandwidth throttling", it is assumed to a neutral network. Vignette 2 allows "zero-rating" but disallows "bandwidth throttling", it is assumed to be the platform-prioritized net. Vignette 3 allows "bandwidth throttling" but disallows "zero-rating", it is assumed to the user-prioritized net. Vignette 4 allows both "zero-rating" and "bandwidth throttling", it is a full prioritized network. These four vignettes are shown in Table 5.1.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Zero-rating</td></tr>
<tr><td>Is NOT allowed</td><td>Is allowed</td></tr>
<tr><td rowspan="2">Bandwidth throttling</td><td>Is NOT allowed</td><td>Vignette 1: Neutral Net</td><td>Vignette 2: Platform-prioritized Net</td></tr>
<tr><td>Is allowed</td><td>Vignette 3: User-prioritized Net</td><td>Vignette 4: Prioritized Net</td></tr>
</table>

Table 5.1 The design of four vignettes

Respondents were asked the following six questions:

1. I am satisfied with my current access rate to the internet.
2. To what extent do you agree that internet service providers should provide zero-rating service?
3. To what extent do you agree that this is an effective strategy to maintain the free-market place of Hong Kong?
4. To what extent do you agree that internet service providers should provide a prioritized service through bandwidth throttling so that customers who pay more get a better access rate to the internet?

5. To what extent do you agree that this is an effective strategy for internet service providers to invest and upgrade their network infrastructures?

6. To what extent do you agree that net neutrality is important to Hong Kong?

The response scale for these questions was a seven-point scale, where 1 = strongly disagree and 7 = strongly agree. According to the statistics, 124 respondents read vignette 1, 128 respondents read vignette 2, 125 respondents read vignette 3 and the remaining 120 respondents read vignette 4, thus all respondents were evenly assigned to 4 vignettes groups.

5.2 Descriptive Results

In order to give a clear and intuitive impression, we show all the polling results by figures in this section. First, the descriptive data on the respondents' use of the Internet is discussed, then the responses to the survey experiment.

Fig. 5.2 shows the network type that respondents use. The result shows that most respondents use both wireless and wireline networks. Note that Wi-Fi is regarded as the wireline network.

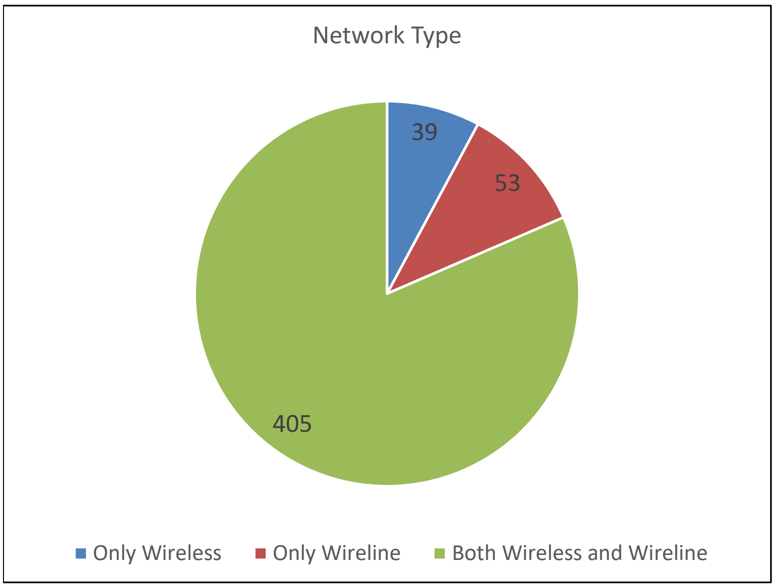


Fig. 5.2 Network type used by respondents

Fig. 5.3 and Fig. 5.4 show the expenditure per month of two types of networks. The majority of the expenditure was found in the range of 0 $\sim$ 200 HKD in both wireline and wireless networks.

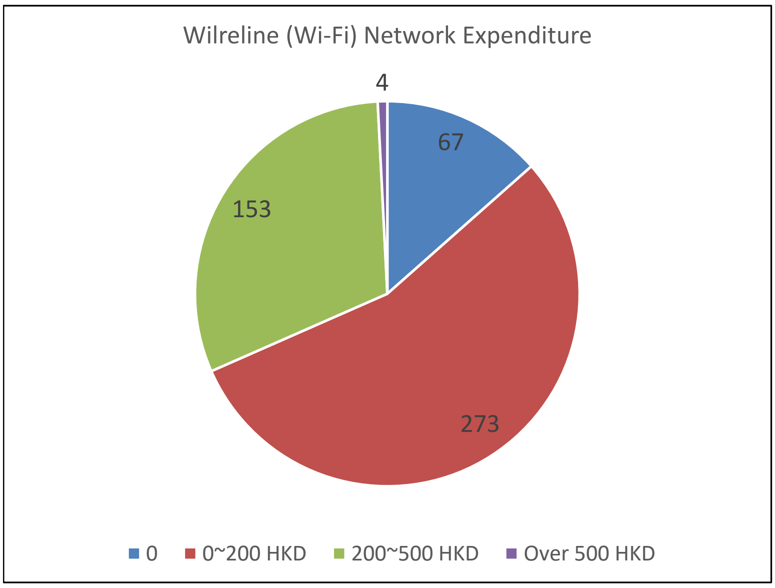


Fig. 5.3 Expenditure of Wireline (Wi-Fi) Network per Month

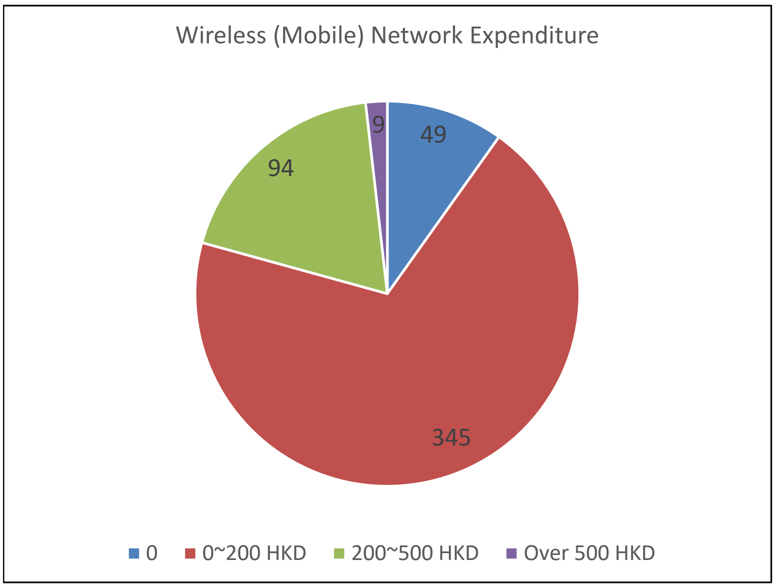


Fig. 5.4 Expenditure of Wireless (Mobile) Network per Month

Fig. 5.5 shows the average time that respondents spent on the Internet. It can be found that 258 of 496 people use the Internet for more than 4 hours a day. The statistics show that 52% of respondents rely heavily on the Internet.

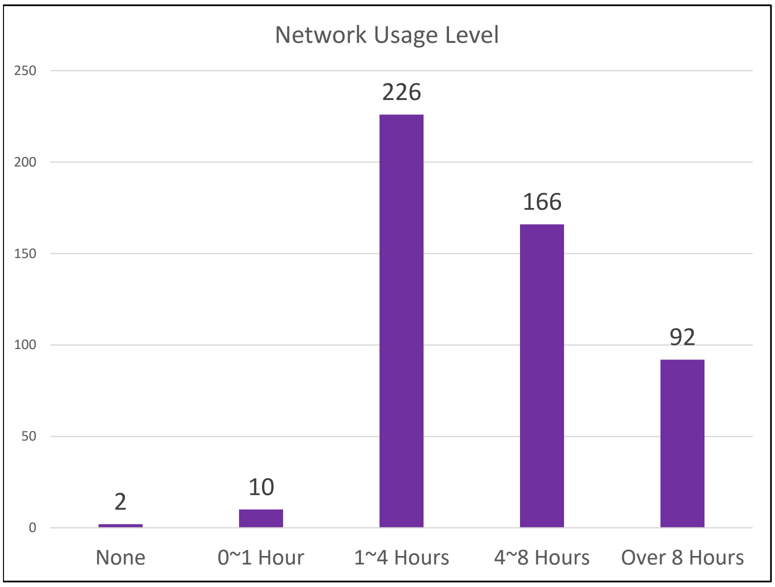


Fig. 5.5 Usage level of respondents

Fig. 5.6 shows the result of respondents' satisfaction with their ISP. The result shows the number of satisfied (235) is greater than the number of dissatisfied (141).

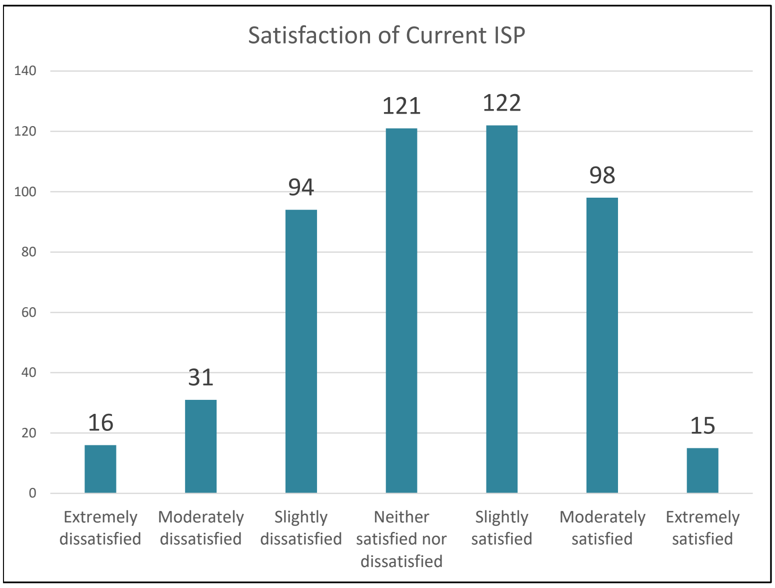
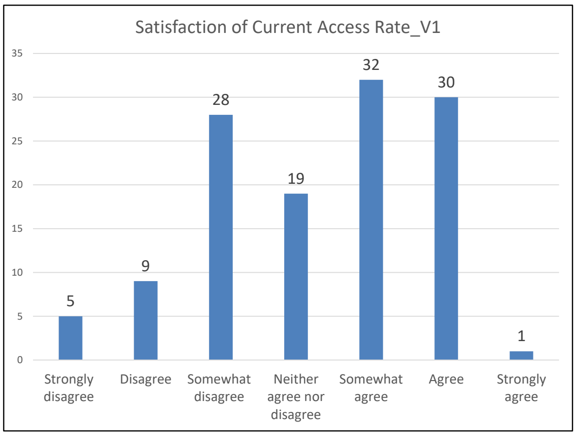
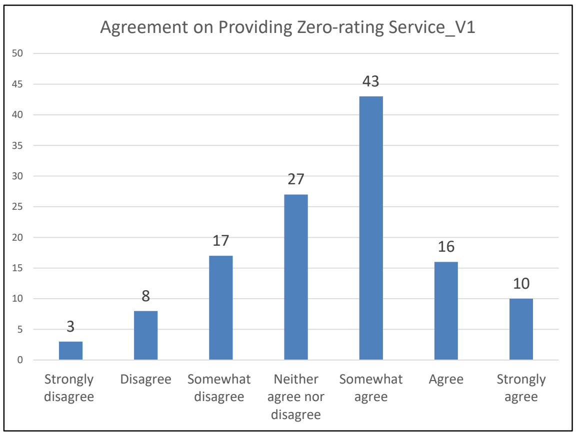


Fig. 5.6 Respondents' satisfaction of current ISP

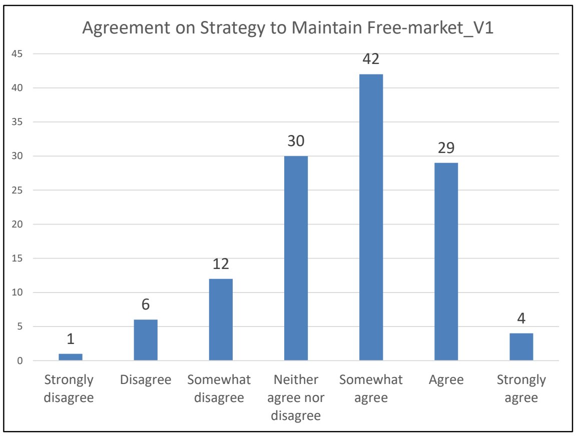
The results of the six questions after four vignettes are given in Figures 5.7 to 5.10, where the accurate statistics of each question in each vignette can be found.



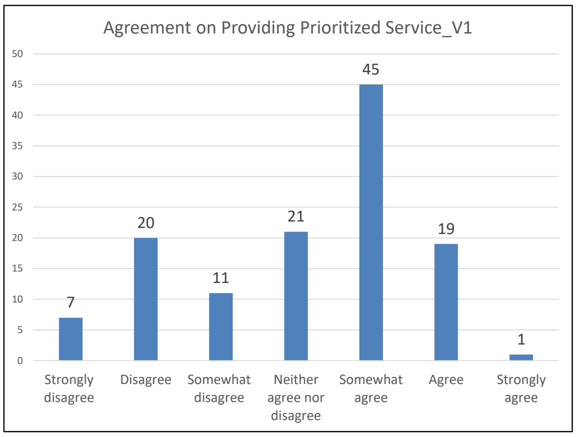
(a) Access rate satisfaction



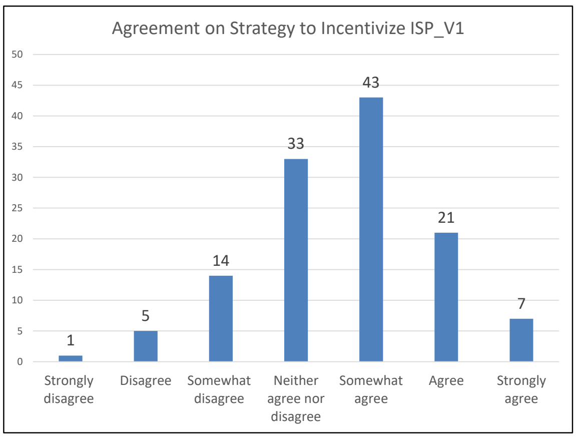
(b) Zero-rating service agreement



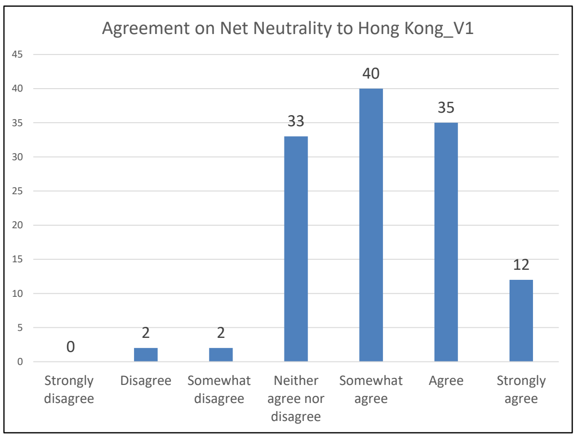
(c) Free-market strategy agreement



(d) Prioritized service agreement

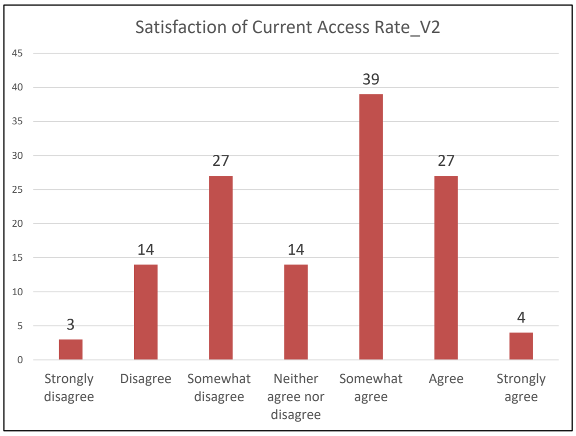


(e) ISP incentive strategy agreement

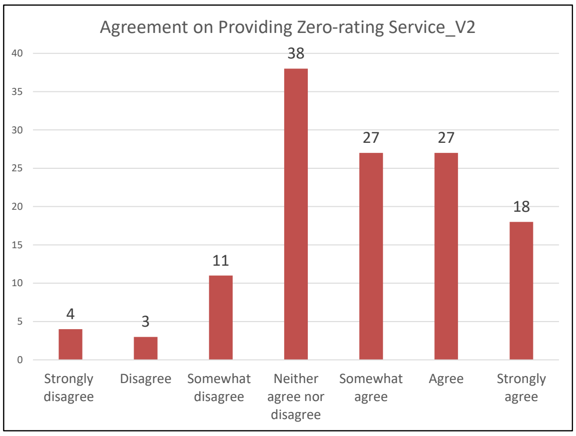


(f) Net neutrality policy agreement

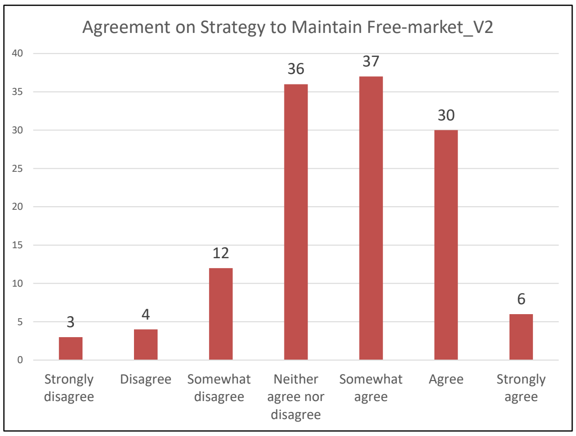
Fig. 5.7 Statistics of vignette 1



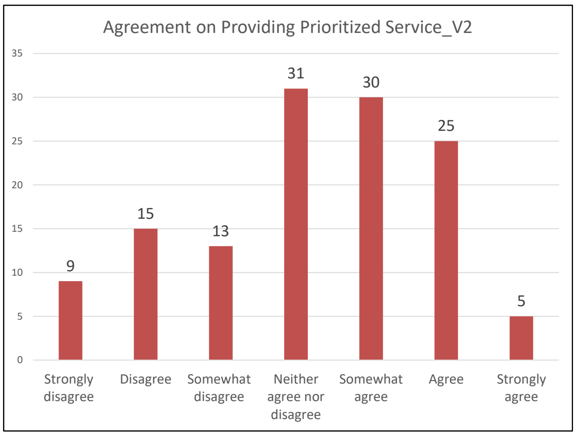
(a) Access rate satisfaction



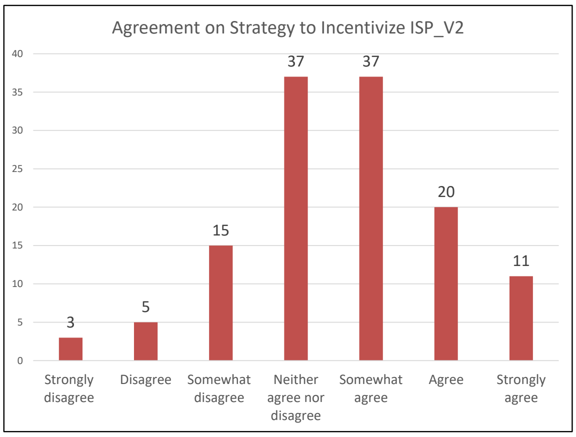
(b) Zero-rating service agreement



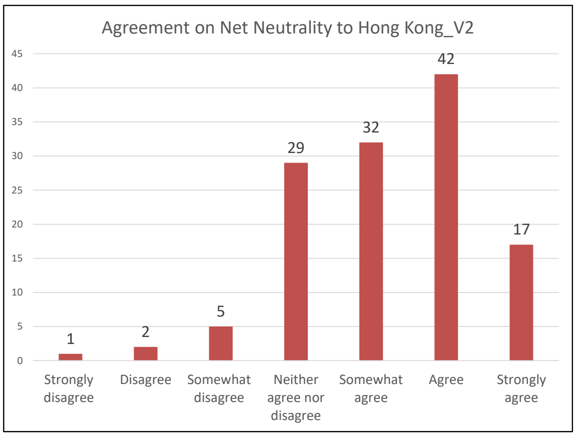
(c) Free-market strategy agreement



(d) Prioritized service agreement

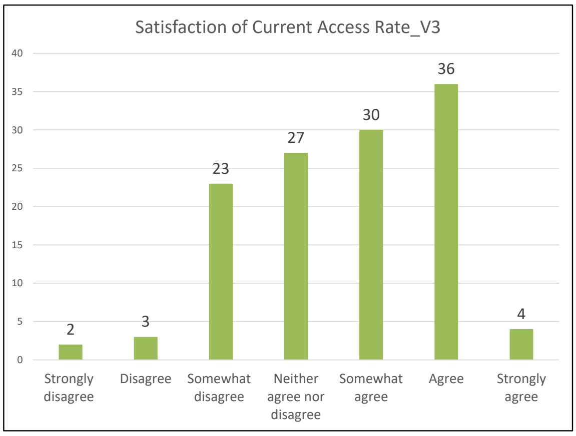


(e) ISP incentive strategy agreement

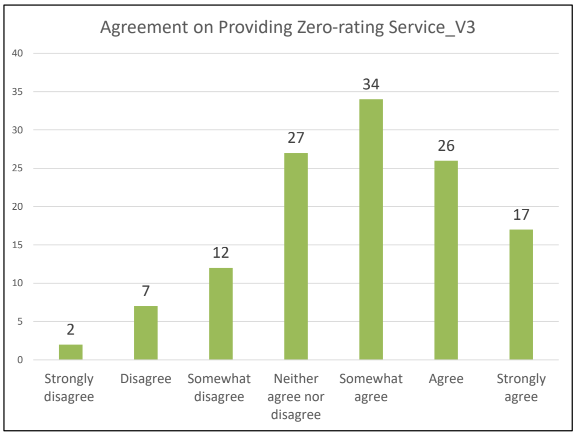


(f) Net neutrality policy agreement

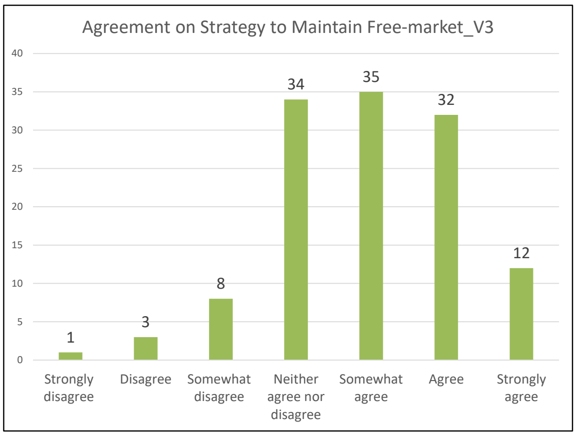
Fig. 5.8 Statistics of vignette 2



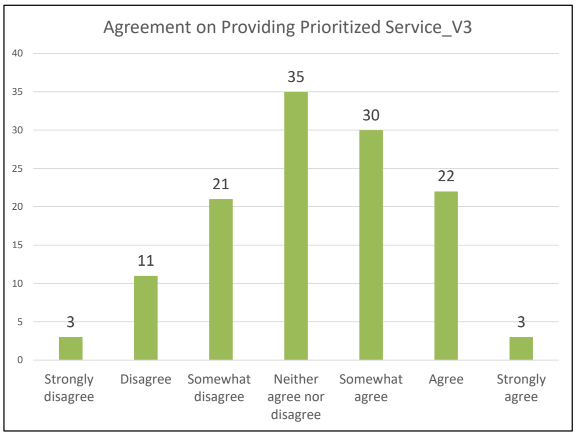
(a) Access rate satisfaction



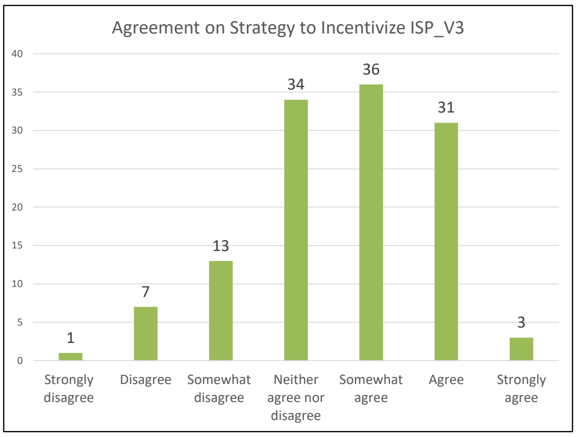
(b) Zero-rating service agreement



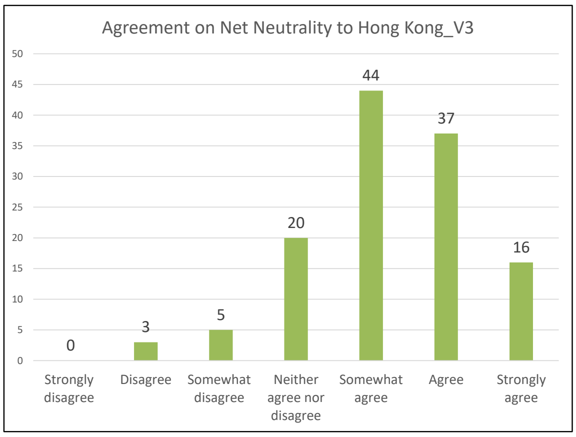
(c) Free-market strategy agreement



(d) Prioritized service agreement

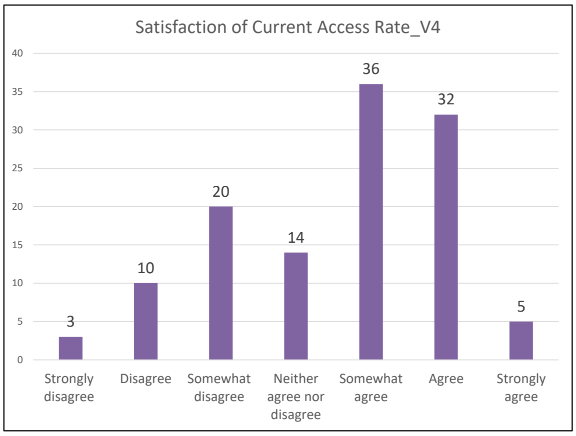


(e) ISP incentive strategy agreement

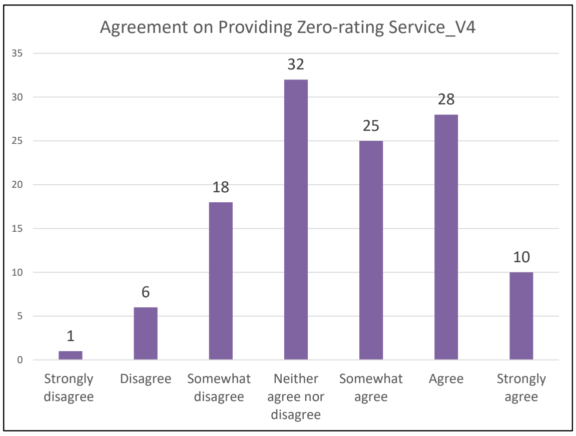


(f) Net neutrality policy agreement

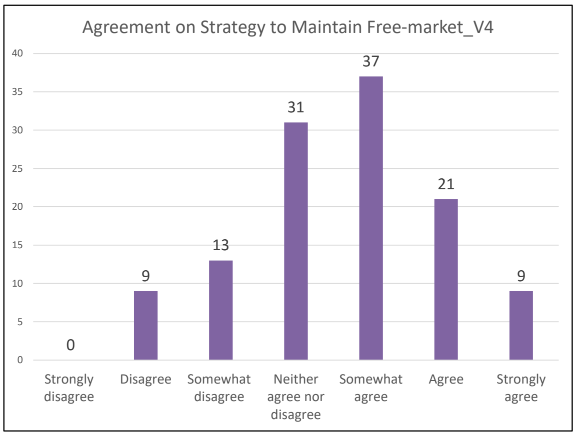
Fig. 5.9 Statistics of vignette 3



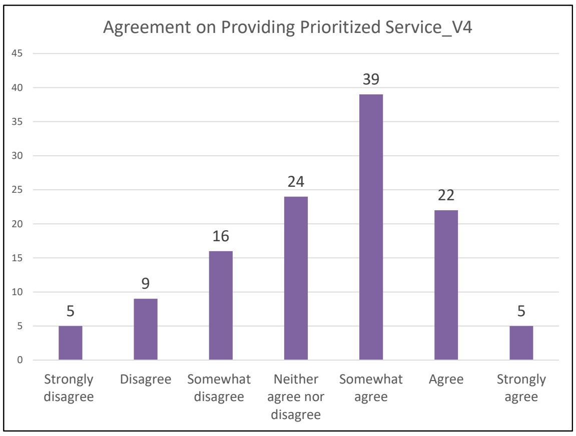
(a) Access rate satisfaction



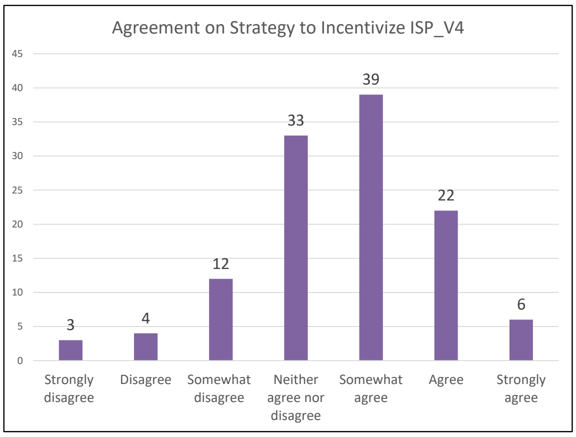
(b) Zero-rating service agreement



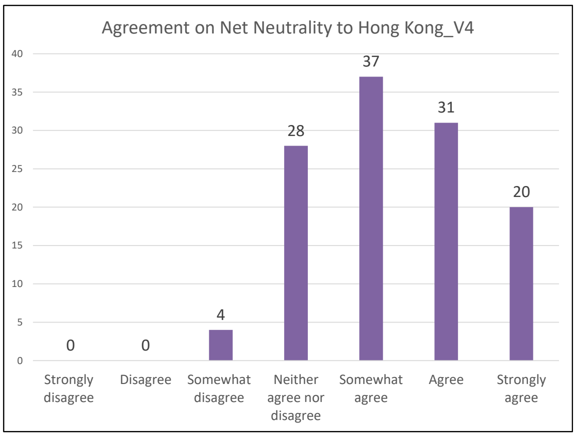
(c) Free-market strategy agreement



(d) Prioritized service agreement



(e) ISP incentive strategy agreement



(f) Net neutrality policy agreement

Fig. 5.10 Statistics of vignette 4

At the end of each vignette, we gave a manipulation check to ensure that respondents understood the topic correctly. Respondents were asked about the definition of Zero-rating. The results are shown in Fig. 5.11. It can be seen from the figure that about 77.26% of the respondents answered correctly.

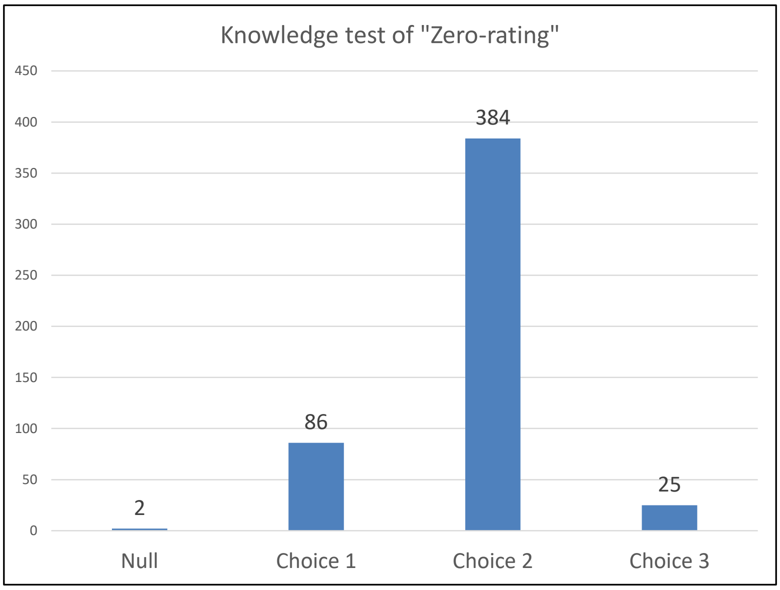


Fig. 5.11 Net neutrality knowledgeability test (Choice 2 is the correct answer, 2 respondents unanswered)

5.3 Analysis

In the analysis part, we adopt the ANOVA (analysis of variance) method. These results are presented in Appendix B, Tables B.1 - B.6, which include the count of participants, the summation of their response rates, the average value, and standard deviation. For the convenience of comparison, we summarize the average and variance of response rates of the six questions in the 4 vignettes, as shown in the following Table 5.2. The maximum and minimum values of the average response rates are highlighted in red and blue, respectively.

	V1		V2		V3		V4	
	Average	Variance	Average	Variance	Average	Variance	Average	Variance
Q1	4.27	2.12	4.32	2.20	4.63	1.72	4.55	2.18
Q2	4.51	1.96	4.83	2.11	4.84	2.10	4.65	1.91
Q3	4.69	1.47	4.67	1.64	4.94	1.52	4.63	1.72
Q4	4.11	2.30	4.20	2.55	4.25	1.85	4.41	2.09
Q5	4.64	1.49	4.59	1.83	4.62	1.53	4.61	1.63
Q6	5.13	1.14	5.21	1.47	5.24	1.30	5.29	1.22

No statistically significant differences

Table 5.2 The summarized results of all respondents' response rate

In order to make the result data of the survey more effective, we screen out the respondents who correctly selected the definition of "zero-rating", performed ANOVA again and summarize these data, as shown in the following Table 5.3. The maximum and minimum values of the average response rates are highlighted as well.

	V1		V2		V3		V4	
	Average	Variance	Average	Variance	Average	Variance	Average	Variance
Q1	4.33	2.29	4.37	2.20	4.68	1.70	4.72	2.10
Q2	4.56	2.25	4.81	2.15	4.89	2.14	4.72	2.10
Q3	4.78	1.44	4.62	1.64	4.97	1.58	4.58	1.88
Q4	4.05	2.72	4.15	2.52	4.23	2.02	4.39	2.20
Q5	4.62	1.59	4.59	1.73	4.60	1.63	4.61	1.57
Q6	5.24	1.15	5.23	1.40	5.26	1.32	5.32	1.22

No statistically significant differences

Table 5.3 The summarized results of "knowledgeable" respondents' response rate

From Table 5.2 and Table 5.3, we observe that there is no statistical significance of the average response rates between 4 vignettes, which means that respondents' answers to these 6 questions were not differentiated by 4 vignettes we have set as anticipated. According to our assumptions, the participants of Vignette 1 will be relatively conservative and therefore agree with the principle of net neutrality. Participants in Vignette 2 tend to give a higher response rate to "zero-rating", participants in Vignette 3 will prefer differentiated services from ISPs, and participants in Vignette 4 will most disagree with the principle of net neutrality.

Specifically, question 1 is the satisfaction of the respondents with their Internet access speed, so the vignettes have little influence on the answer to this issue. Interestingly, it can be found that the highest agreement on the "zero-rating" policy appears in vignette 3, where "zero-rating" is banned but "bandwidth throttling" is advocated. Such agreement supposed to be found in vignette 2. For the agreement on free-market strategy, the highest agreement can be found in vignette 3, which is consistent with the assumption of vignette 3: high efficiency in a fair competition environment. As for the agreement on prioritized service, respondents from vignette 4, which allows both "zero-rating" and "bandwidth throttling", gave the highest agreement. Unexpectedly, the respondents of Vignette 1 gave the highest agreement on the strategy that can stimulate ISPs to upgrade network infrastructure, where vignette 1 itself is the most conservative and considered to be neutral. In question 6, it is also surprising that the respondents of vignette 4 valued the importance of net neutrality to Hong Kong most, whereas vignette 4 was set as far from net neutral.

5.4 Conclusions and Summary

This survey experiment was carried on two dimensions: equity and efficiency. For equity, we used the term "zero-rating", which describes certain content providers subsidize their subscribers, to test respondents' sensitivity to fair competition. For efficiency, we used the term “bandwidth-throttling”, which describes ISPs prioritizing different users based on the price they paid, to test respondents’ attitudes of the Internet’s efficiency.

From the descriptive data collected before the vignette questions, we found that most participants use both mobile and wired networks. For both wired and wireless networks, the monthly charges fall in the range of $0 \sim 200$ Hong Kong dollars. Most Hong Kong residents are satisfied with their Internet access speed. The ANOVA analysis of the survey experiments found no statistical differences between the four vignette treatments. From the perspective of improvement to the study design, we offer three approaches:

1. The complexity of the topic. Compared with other policy topics that are often used in public policy research, the issue of net neutrality and related concepts are very complex and highly technical, and not so straightforward to understand. Respondents needed to read a total of 4 pages of terminology explanations before they can start answering the questionnaire. Many respondents did, however, understand these complete terms: at the end of the questionnaire, 77.26% of the respondents chose the correct answer to answer the definition of “zero-rating” in the manipulation check.

2. Sample size. In public surveys, 1,000 samples are usually considered sufficient to ensure representation when random sampling is used. This study employed a convenience sample, 500 valid responses across the 4 vignettes after three months’ in the field. The sample size may be too small to obtain statistical differences between the vignettes. Further, respondents received remuneration of HK$50 after answering all the questions, but as mentioned in the first aspect, the complexity of the questions may have caused some respondents to give up answering halfway or just rush to answer to get remuneration.

3. The design of the questionnaire. We have not provided the respondents with specific quantitative information in each vignette but only use general language to describe

the different results. The quantitative description gives a more intuitive feeling, such as how many megabits per second of bandwidth can be enjoyed at a certain level of the monthly payment, and how many seconds of the buffering time can be reduced by paying more money to the ISP. This may be the reason that there is no statistical significance of the vignette questions in the ANOVA analysis. Another potential reason may be the familiarity of Hong Kong citizens with various commercial offers and promotions, which led to some respondents still choosing to support "zero-rating" on the premise that they clearly understand "zero-rating" may violate the net neutrality principle.

Based on the above three aspects, we will further simplify the complexity of the questionnaire in future surveys, collect more respondents, and provide more quantitative information so that they can make intuitive decisions, and design vignettes that close to the real business practice.

We have conducted the first survey experiment on the net neutrality policy in Hong Kong. Although the results did not show the statistical significance as we expected, we have gained useful experience to explain complex concepts to the public. The participants' response of the net neutrality importance to Hong Kong motivates us to carry on such research in a better way.

Chapter 6

Conclusion

6.1 General Conclusion

This thesis studies the topic of net neutrality. We gave a comprehensive review of the literature from three perspectives, namely policy, economics, and engineering. The concept of net neutrality has been controversial since the day it was proposed. Different stakeholders have different interpretations of this emerging term: Network providers believe that they build and maintain the network communication facilities, so that they deserve a corresponding return. Content providers believe that they provide rich content which prospered the network community, and one of the prerequisites of this prosperity is that the network treats all data equally. The development of technology makes the end-users of the network hope to enjoy more content at a lower price. Network development is unbalanced across different regions. A relatively general conclusion can be made: Network operators are concerned whether they can profit from the network facilities they invest in, and are opposed to network neutrality. On the other hand, content providers and end-users often stand on the side of network neutrality.

Network neutrality is a complex concept, the key concept of which is whether to allow differentiated services (or paid priority) to exist in the network. However, by looking back at the historical record and following the successive refinements of policy research on network neutrality, we found that certain content providers and network operators overlap in their roles, as manifested in the concept of vertical integration, i.e., a network operator not only provides network connection services to end-users, but also provides them with content. Furthermore, in less developed countries and regions, some content providers have signed agreements with local network operators to subsidize their users by means of "zero-rate" (i.e., free of charge or compensation). These two measures violate the fair competition principle of the market. Researchers in the economic and engineering fields pay more attention to the investment incentives of ISPs and the social welfare of the entire society. Most of their research points out that maintaining market competition is more important than the concept of net neutrality. In addition to the detailed literature review in Chapter 2, we also discuss in Chapter 3 the concept of QoS in data networks and review the process of evolution of the architecture for implementing differentiated services in computer networks. More importantly, we explore the specific implementation of differentiated services in each layer of the OSI7 layer model from a multi-layered perspective.

By selecting different queuing models, in Chapter 4, we have studied the impact of differentiated services from the perspective of network operators and end-users. Unlike traditional telephone networks, data networks do not reject new customers to receive services during critical loading. However, the newly arrived customers will affect the QoE of the existing users, which is called "soft blocking". Considering the strong relation between soft blocking and hard blocking in the telephone network, we borrowed the Erlang B formula to measure the impact of introducing differentiated services on the cost incurred by ISPs. Next, we have evaluated the impact of differentiated services on the end-users' experience with a Processor Sharing queueing model, which focuses mainly on the delay. The simulation results show that distributing bandwidth to all users uniformly is not an optimal solution to meet users' heterogeneous demands. Transmitting data through various protocols including TCP and UDP has been accepted by the public for a long time and it has not been criticized by net neutrality proponents, such practice can be regarded as a social norm. In fact, due to the use of different transmission protocols, data packets in the network are not treated “equally”, therefore we believe that the statement in the net neutrality principle “treat all the packets equally” is potentially too strong. Based on this fact, we proposed to prioritize data traffic according to the transmission protocols. We evaluated this proposal through real traffic datasets. We have demonstrated that such an approach can improve the users’ service quality and address the network neutrality dispute in the future.

The Internet is playing an increasingly important role in today’s human society. Hong Kong has yet to legislate or give relevant guidance on issues related to net neutrality. It is an indisputable fact that Hong Kong’s infrastructure of communication networks is unbalanced. The lack of a reasonable network policy will not motivate ISPs to invest in remote areas. In the long run, it will hurt Hong Kong’s competitiveness. To this end, we started a survey on net neutrality for end-users in Hong Kong in November 2018. The results of the survey, listed in Chapter 5, found that people are willing to accept “zero-rating” services and “bandwidth throttling” (or paid prioritization) while still claim the importance of the net neutrality principle to Hong Kong, therefore the results did not show the statistical significance. The reason for this may be due to the relative complexity of network neutrality and related concepts, and we did not provide the respondents with quantitative information in the survey. In any case, this is the first investigative experiment on net neutrality policy in Hong Kong. Although the results of the survey did not show the statistical significance as we expected, the participants’ response of the net neutrality importance to Hong Kong motivates us to carry on such research.

6.2 Future work

This thesis focuses on the last mile of Internet services. After a comprehensive review and exploratory research on the implementation of priority services on the last mile of the Internet, we summarize the limitation of our work and outline the work to be carried out in the future:

- In the numerical part, we use queuing theory as the main tool for analyzing QoS and QoE, such as blocking probability and network delay, which clearly show the benefits of applying differentiated services. However, the queuing models we use are not completely applicable to data networks. More accurate queueing models that accurately capture data network behaviour could be developed for the purpose of assessing service differentiation aspects in data networks.

- The second point is that we only consider classic broadband (including mobile) network service scenarios where content providers send their data to end-users through service providers' lane. With the popularization of smartphones, more people are willing to share their activity in real-time using live video stream services including Facebook Live and YouTube Live. This lead to a new ecosystem of Internet services. Some users simultaneously play the service provider and consumer roles. Little research has been done on the impact this has on the Internet's ecosystem.

- In this work, one of our contributions is that we conducted public polling to hear the opinion of the public. However, this survey is one way. The other direction of the investigation is to listen to the opinions of service providers and content providers and find out their concerns. We believe that such a two-way survey will form a complete picture, and the following-up findings and conclusions will be more instructive in network management policymaking.

References

[1] J. M. Bauer, "Dynamic effects of network neutrality," in *Proceedings of 35th Research Conference on Communication, Information and Internet Policy*, Alexandria, VA, Sep./Oct. 2006.

[2] B. E. Hermalin and M. L. Katz, "The economics of product-line restrictions with an application to the network neutrality debate," *Information Economics and Policy*, vol. 19, no. 2, pp. 215–248, Jun. 2007.

[3] J. Musacchio, G. Schwartz, and J. Walrand, "A two-sided market analysis of provider investment incentives with an application to the net-neutrality issue," *Review of Network Economics*, vol. 8, no. 1, Mar. 2009.

[4] J. Walrand, "Economic models of communication networks," *Performance Modeling and Engineering*, pp. 57–89, 2008.

[5] G. Kramer, B. Mukherjee, S. Dixit, Y. Ye, and R. Hirth, "Supporting differentiated classes of service in ethernet passive optical networks," *Journal of Optical Networking*, vol. 1, no. 8, pp. 280–298, Aug./Sep. 2002.

[6] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the differentiated services field (DS Field) in the IPv4 and IPv6 headers," RFC2474, Dec. 1998. [Online]. Available: https://tools.ietf.org/html/rfc2474

[7] C. T. Marsden, "Comparative case studies in implementing net neutrality: a critical analysis of zero rating," *SCRIPTed*, vol. 13, pp. 1–39, May 2016.

[8] R. Gibbens, R. Mason, and R. Steinberg, "Internet service classes under competition," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2490–2498, Dec. 2000.

[9] "The average web page has almost doubled in size since 2010," Jun. 2013. [Online]. Available: http://www.webperformancetoday.com/2013/06/05/web-page-growth-2010-2013/

[10] Cisco, "Cisco visual networking index: Forecast and methodology, 2016-2021," Sep. 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[11] T. Wu, "Network neutrality, broadband discrimination," *Journal on Telecommunications & High Technology Law*, vol. 2, pp. 141–176, 2003.

[12] V. Cerf, "The open internet what it is, and why it matters," *Telecommunications Journal of Australia*, vol. 59, no. 2, pp. 18–1, 2009.

[13] V. Misra, "Routing money, not packets," *Communications of the ACM*, vol. 58, no. 6, pp. 24–27, Jun. 2015.

[14] K. Werbach, "Breaking the ice: Rethinking telecommunications law for the digital age," *Journal on Telecommunications and High Technology Law*, vol. 4, p. 59, 2005.

[15] A. Banerjee and C. M. Dippon, "Communications regulation and policy under convergence: advancing the state of the debate," in *Proc. 16th 2006 Biennial Conference ITS*, Beijing, China, Jun. 2006.

[16] T. Wu, "The broadband debate, a user's guide," *Journal on Telecommunications & High Technology Law*, vol. 3, pp. 69–96, 2004.

[17] D. S. Isenberg, "The four 'net freedoms in chairman powell's own words," Aug. 2005. [Online]. Available: http://isen.com/blog/2005/08/four-net-freedoms-in-chairman-powells.html

[18] P. Crocioni, "Net neutrality in europe: Desperately seeking a market failure," *Telecommunications Policy*, vol. 35, no. 1, pp. 1–11, Jan. 2011.

[19] B. Van Schewick, "Towards an economic framework for network neutrality regulation," *Journal on Telecommunications & High Technology Law*, vol. 5, pp. 329–392, 2006.

[20] D. Mccullagh, "Telco agrees to stop blocking VoIP calls," Mar. 2005. [Online]. Available: https://www.cnet.com/news/telco-agrees-to-stop-blocking-voip-calls/

[21] T. Karr, "Net blocking: A problem in need of a solution," Aug. 2014. [Online]. Available: https://www.huffingtonpost.com/timothy-karr/net-blocking-a-problem-in_b_5695997.html

[22] D. Mccullagh, "FCC formally rules comcast's throttling of bittorrent was illegal," Aug. 2008. [Online]. Available: https://www.cnet.com/news/fcc-formally-rules-comcasts-throttling-of-bittorrent-was-illegal/

[23] A. Frank, "The FCC, net neutrality, and the principle of transparency," Sep. 2009. [Online]. Available: https://blogs.gartner.com/andrew_frank/2009/09/22/the-fcc-net-neutrality-and-the-principle-of-transparency/

[24] E. Wyatt, "F.C.C. begins investigation into quality of Internet download speeds," May 2014. [Online]. Available: https://www.nytimes.com/2014/06/14/business/media/FCC-inquiry-into-ties-between-content-companies-and-service-providers.html?smid=pl-share

[25] J. Brodkin, "Verizon accused of throttling Netflix and YouTube, admits to "video optimization"," Jul. 2017. [Online]. Available: https://www.huffingtonpost.com/timothy-karr/net-blocking-a-problem-in_b_5695997.html

[26] J. Krämer, L. Wiewiorra, and C. Weinhardt, "Net neutrality: A progress report," *Telecommunications Policy*, vol. 37, no. 9, pp. 794–813, Oct. 2013.

[27] E. Rosenberg, "What you should know about net neutrality," May 2018. [Online]. Available: https://www.investopedia.com/articles/investing/033015/what-you-should-know-about-net-neutrality.asp

[28] T. Garrett, L. E. Setenareski, L. M. Peres, L. C. Bona, and E. P. Duarte, "Monitoring network neutrality: A survey on traffic differentiation detection," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2486–2517, 2018.

[29] L. Lessig, "Testimony of larry lessig, hearing on network neutrality," Senate Committee on Commerce, Science and Transportation, Feb. 2006.

[30] B. D. Herman, "Opening bottlenecks: On behalf of mandated network neutrality," *Federal communications law journal*, vol. 59, pp. 107–160, 2006.

[31] J. M. Peha, "The benefits and risks of mandating network neutrality, and the quest for a balanced policy," in *Proceedings of 34th Telecommunications Policy Research Conference*, Washington DC, Sep./Oct. 2006, pp. 1–23.

[32] G. S. Ford, T. Koutsky, and L. J. Spiwak, "The efficiency risk of network neutrality rules," *Phoenix Center Policy Bulletin*, no. 16, May 2006.

[33] S. Jordan, "A layered network approach to net neutrality," *International Journal of Communication*, vol. 1, pp. 427–460, 2007.

[34] D. D. Clark, "Network neutrality: Words of power and 800-pound gorillas," *International Journal of Communication*, vol. 1, pp. 701–708, 2007. [Online]. Available: http://ijoc.org/index.php/ijoc/article/download/158/83

[35] H. K. Cheng, S. Bandyopadhyay, and H. Guo, "The debate on net neutrality: A policy perspective," *Information systems research*, vol. 22, no. 1, pp. 60–82, Mar. 2011.

[36] P. Maillé, K. Pires, G. Simon, and B. Tuffin, "How neutral is a CDN? an economic approach," in *Proceedings of 10th International Conference on Network and Service Management (CNSM) and Workshop*, Rio de Janeiro, Brazil, Nov. 2014, pp. 336–339.

[37] G. R. Faulhaber, "Network neutrality: The debate evolves," *International Journal of Communication*, vol. 1, pp. 680–700, 2007.

[38] P. Ganley and B. Allgrove, "Net neutrality: A user's guide," *Computer Law & Security Review*, vol. 22, no. 6, pp. 454–463, 2006.

[39] C. Kang, "F.C.C. repeals net neutrality rules," Dec. 2017. [Online]. Available: https://www.nytimes.com/2017/12/14/technology/net-neutrality-repeal-vote.html

[40] N. Feamster, "New jersey takes up net neutrality: A summary, and my experiences as a witness," Mar. 2018. [Online]. Available: https://freedom-to-tinker.com/2018/03/14/new-jersey-takes-up-net-neutrality-a-summary-and-my-experiences-as-a-witness/

[41] S. Wong, J. Rojas-Mora, and E. Altman, "Public consultations on net neutrality 2010: USA, EU and France," in *Proc. of NetCoop 2010*, Ghent, Belguim, Nov. 2010.

[42] J. M. Peha, W. H. Lehr, and S. Wilkie, "The state of the debate on network neutrality," *International Journal of Communication*, vol. 1, pp. 709–716, Jul. 2007.

[43] G. Goth, “The global net neutrality debate: Back to square one?” *IEEE Internet Computing*, vol. 14, no. 4, pp. 7–9, Jul./Aug. 2010.

[44] A. Cooper, “How regulation and competition influence discrimination in broadband traffic management: A comparative study of net neutrality in the united states and the united kingdom,” Ph.D. dissertation, University of Oxford, Sep. 2013.

[45] C. S. Yoo, “Beyond network neutrality,” *Harvard Journal of Law & Technology*, vol. 19, no. 1, 2005.

[46] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, “A quantitative measure of fairness and discrimination,” *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 1984.

[47] F. P. Kelly, A. K. Maulloo, and D. K. Tan, “Rate control for communication networks: shadow prices, proportional fairness and stability,” *Journal of the Operational Research society*, vol. 49, no. 3, pp. 237–252, 1998.

[48] L. Massoulié and J. Roberts, “Bandwidth sharing: objectives and algorithms,” in *Proc. IEEE INFOCOM’99*, vol. 3, New York, NY, Mar. 1999, pp. 1395–1403.

[49] M. Zukerman, L. Tan, H. Wang, and I. Ouveysi, “Efficiency-fairness tradeoff in telecommunications networks,” *IEEE Communications Letters*, vol. 9, no. 7, pp. 643–645, Jul. 2005.

[50] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, “A game theoretic framework for bandwidth allocation and pricing in broadband networks,” *IEEE/ACM transactions on networking*, vol. 8, no. 5, pp. 667–678, Oct. 2000.

[51] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” *IEEE/ACM Transactions on networking*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[52] F. Kelly *et al.*, “Fairness and stability of end-to-end congestion control,” *European journal of control*, vol. 9, no. 2-3, pp. 159–176, 2003.

[53] M. R. Salvador, M. M. Uesono, and N. L. da Fonseca, “A packet ring fairness protocol and its impact on TCP fairness,” in *Proc. GLOBECOM’05*, vol. 2, St. Louis, MO, Nov./Dec. 2005.

[54] M. J. Neely, E. Modiano, and C.-P. Li, “Fairness and optimal stochastic control for heterogeneous networks,” *IEEE/ACM Transactions On Networking*, vol. 16, no. 2, pp. 396–409, Apr. 2008.

[55] Z. Rosberg, J. Matthews, and M. Zukerman, “A network rate management protocol with TCP congestion control and fairness for all,” *Computer Networks*, vol. 54, no. 9, pp. 1358–1374, Jun. 2010.

[56] C. S. Yoo, “Network neutrality and the economics of congestion,” *The Georgetown Law Journal*, vol. 94, pp. 1847–1908, 2005.

[57] J. Pil Choi and B.-C. Kim, “Net neutrality and investment incentives,” *The RAND Journal of Economics*, vol. 41, no. 3, pp. 446–471, Aug. 2010.

[58] A. Thierer, "Are "dumb pipe" mandates smart public policy? vertical integration, net neutrality, and the network layers model," in *Net neutrality or net neutering: Should broadband Internet services be regulated*. Springer, 2006, pp. 73–108.

[59] R. S. Lee and T. Wu, "Subsidizing creativity through network design: Zero-pricing and net neutrality," *The Journal of Economic Perspectives*, vol. 23, no. 3, pp. 61–76, 2009.

[60] C. S. Hemphill, "Network neutrality and the false promise of zero-price regulation," *Yale Journal on Regulation*, vol. 25, no. 2, pp. 135–179, 2008.

[61] D. L. Weisman and R. B. Kulick, "Price discrimination, two-sided markets, and net neutrality regulation," *Tulane Journal of Technology and Intellectual Property*, vol. 13, pp. 81–102, 2010.

[62] R. J. Deneckere and R. Preston McAfee, "Damaged goods," *Journal of Economics & Management Strategy*, vol. 5, no. 2, pp. 149–174, Jun. 1996.

[63] H. H. Gharakheili, A. Vishwanath, and V. Sivaraman, "Pricing user-sanctioned dynamic fast-lanes driven by content providers," in *Proc. IEEE Conference on Computer Communications Workshops 2015 (INFOCOM WKSHPS)*, Hong Kong, China, Apr./May 2015, pp. 528–533.

[64] M. P. Shane Greenstein and T. Valletti, "Net neutrality: A fast lane to understanding the trade-offs," *Journal of Economic Perspectives*, vol. 30, no. 2, pp. 127–150, 2016.

[65] "For zero-rated deals, OTT providers can no longer assume the carrier will pay," Feb. 2014. [Online]. Available: https://www.fiercewireless.com/europe/for-zero-rated-deals-ott-providers-can-no-longer-assume-carrier-will-pay

[66] F. Marini Balestra and R. Tremolada, "The EU debate on net neutrality: What about zero-rating?" *Computer and Telecommunications Law Review*, vol. 21, no. 5, 2015.

[67] C. T. Marsden, *Zero Rating and Mobile Net Neutrality*. Cham: Springer International Publishing, 2016, pp. 241–260.

[68] A. M. Kakhki, F. Li, D. Choffnes, E. Katz-Bassett, and A. Mislove, "BingeOn under the microscope: Understanding T-mobiles zero-rating implementation," in *Proceedings of the 2016 workshop on QoE-based Analysis and Management of Data Communication Networks*, Florianopolis , Brazil, Aug. 2016, pp. 43–48.

[69] M. Asghari, S. Yousefi, and D. Niyato, "A mobile network operator's decision for partnership with zero-rating Internet platforms," in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017.

[70] J. Kastrenakes, "A timeline of Netflix's conflicting stances on net neutrality," Mar. 2017. [Online]. Available: https://www.theverge.com/2017/3/20/14960154/netflix-net-neutrality-stances-timeline

[71] A. Moshirnia, "Zero-rating: Price discrimination in an era of net neutrality," 2015. [Online]. Available: https://www.americanbar.org/publications/infrastructure/2014-15/summer/zerorating_price_discrimination_an_era_net_neutrality.html

[72] "Free basics platform." [Online]. Available: https://info.internet.org/en/story/platform/

[73] "Wikipedia zero." [Online]. Available: https://wikimediafoundation.org/wiki/Wikipedia_Zero

[74] R. B. Ferreira, P. Brancher, C. T. R. da Silva, and A. S. Advogados, "Communications: regulation and outsourcing in brazil: overview." [Online]. Available: https://uk.practicallaw.thomsonreuters.com/8-619-3338?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1

[75] N. Economides, "Net neutrality, non-discrimination and digital distribution of content through the Internet," *I/S: A Journal of Law and Policy for the Information Society*, vol. 4, no. 2, pp. 209–233, 2008.

[76] F. Schuett, "Network neutrality: A survey of the economic literature," *Review of Network Economics*, 2010.

[77] A. Lafarre, "Net neutrality and price discrimination: Using skype as an "arbitrage" opportunity," Master's thesis, Tilburg University, Jul. 2014.

[78] R. W. Hahn and S. Wallsten, "The economics of net neutrality," *The Economists' Voice*, vol. 3, no. 6, pp. 1–7, Jun. 2006.

[79] R. T. B. Ma, D.-M. Chiu, J. Lui, V. Misra, and D. Rubenstein, "Internet economics: The use of shapley value for ISP settlement," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 3, pp. 775–787, Jun. 2010.

[80] R. T. B. Ma, D.-M. Chiu, J. C. Lui, V. Misra, and D. Rubenstein, "On cooperative settlement between content, transit, and eyeball Internet service providers," *IEEE/ACM Transactions on networking*, vol. 19, no. 3, pp. 802–815, Jun. 2011.

[81] K. G. Wilson, "The last mile: Service tiers versus infrastructure development and the debate on Internet neutrality," *Canadian Journal of Communication*, vol. 33, no. 1, 2008.

[82] *Broadband Internet performance: a view from the gateway*, Toronto, Ontario, Canada, Aug. 2011.

[83] *Accelerating last-mile web performance with popularity-based prefetching*, Helsinki, Finland, Aug. 2012.

[84] S. Sundaresan, N. Feamster, and R. Teixeira, "Home network or access link? locating last-mile downstream throughput bottlenecks," in *Proc. PAM 2016 - Passive and Active Measurement Conference*, Heraklion, Greece, Mar. 2016, pp. 111–123.

[85] J. K. MacKie-Mason and H. R. Varian, "Pricing congestible network resources," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1141–1149, Sep. 1995.

[86] R. J. Gibbens, R. A. Mason, and R. Steinberg, "Multiproduct competition between congestible networks," Working Paper WP 26/98, Judge Inst. Manage. Studies, Univ. Cambridge, Sep. 1998.

[87] M. A. Jamison and J. A. Hauge, “Getting what you pay for: Analyzing the net neutrality debate,” *Mimeo, Social Science Research Network*, Apr. 2008.

[88] J. Musacchio and S. Wu, “The price of anarchy in competing differentiated services networks,” in *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2008, pp. 615–622.

[89] E. Altman, P. Bernhard, S. Caron, G. Kesidis, J. Rojas-Mora, and S. Wong, “A study of non-neutral networks with usage-based prices,” in *Proc. 3rd Workshop on Economic Traffic Management*, Amsterdam, Netherlands, Sep. 2010.

[90] S. Caron, G. Kesidis, and E. Altman, “Application neutrality and a paradox of side payments,” in *Proceedings of the Re-Architecting the Internet Workshop*, Philadelphia, Nov. 2010, p. 9.

[91] Z.-L. Zhang, P. Nabipay, A. Odlyzko, and R. Guerin, “Interactions, competition and innovation in a service-oriented Internet: An economic model,” in *Proc. IEEE INFOCOM 2010*, San Diego, CA, Mar. 2010, pp. 1–5.

[92] N. Economides and B. E. Hermalin, “The economics of network neutrality,” *The RAND Journal of Economics*, vol. 43, no. 4, pp. 602–629, Dec. 2012.

[93] P. Njoroge, A. Ozdaglar, N. E. Stier-Moses, and G. Y. Weintraub, “Investment in two-sided markets and the net neutrality debate,” *Review of Network Economics*, vol. 12, no. 4, pp. 355–402, 2013.

[94] M. Bourreau, F. Kourandi, and T. Valletti, “Net neutrality with competing Internet platforms,” *The Journal of Industrial Economics*, vol. 63, no. 1, pp. 30–73, Mar. 2015.

[95] A. Odlyzko, “Paris metro pricing for the Internet,” in *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 140–147.

[96] C.-K. Chau, Q. Wang, and D.-M. Chiu, “On the viability of paris metro pricing for communication and service networks,” in *Proceedings of IEEE INFOCOM 2010*, San Diego, CA, Mar. 2010, pp. 1–9.

[97] M. Rysman, “The economics of two-sided markets,” *Journal of economic perspectives*, vol. 23, no. 3, pp. 125–143, 2009.

[98] E. Altman, A. Legout, and Y. Xu, “Network non-neutrality debate: An economic analysis,” in *Proceedings of 10th International IFIP TC6 Networking Conference*, Valencia, Spain, May 2011, pp. 68–81.

[99] N. Economides and J. Tåg, “Network neutrality on the Internet: A two-sided market analysis,” *Information Economics and Policy*, vol. 24, no. 2, pp. 91–104, 2012.

[100] M. Yacoubi, M. Emelianenko, and N. Gautam, “Pricing in next generation networks: a queuing model to guarantee QoS,” *Performance Evaluation*, vol. 52, no. 1, pp. 59–84, Mar. 2003.

[101] X. Wang, R. T. B. Ma, and Y. Xu, “On optimal two-sided pricing of congested networks,” *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 1, no. 1, pp. 6:1–6:28, Jun. 2017.

[102] J. Krämer and L. Wiewiorra, "Network neutrality and congestion sensitive content providers: Implications for content variety, broadband investment, and regulation," *Information Systems Research*, vol. 23, no. 4, pp. 1303–1321, May 2012.

[103] R. T. B. Ma and V. Misra, "Congestion and its role in network equilibrium," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2180–2189, Dec. 2012.

[104] R. T. B. Ma, D.-m. Chiu, J. C. S. Lui, V. Misra, and D. Rubenstein, "Internet economics: The use of shapley value for ISP settlement," in *Proceedings of the 2007 ACM CoNEXT*, New York, NY, Dec. 2007.

[105] R. T. B. Ma and V. Misra, "The public option: A nonregulatory alternative to network neutrality," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1866–1879, Dec. 2013.

[106] J. Tang and R. T. B. Ma, "Regulating monopolistic ISPs without neutrality," in *Proceedings of IEEE 22nd International Conference on Network Protocols*, The Research Triangle, NC, Oct. 2014, pp. 374–384.

[107] J. Wang, R. T. B. Ma, and D. M. Chiu, "Paid prioritization and its impact on net neutrality," in *Proc. 2014 IFIP Networking Conference*, Trondheim, Norway, Jun. 2014, pp. 1–9.

[108] R. T. B. Ma, "Usage-based pricing and competition in congestible network service markets," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 3084–3097, Oct. 2016.

[109] ——, "Subsidization competition: Vitalizing the neutral Internet," *IEEE/ACM Transactions on Networking*, no. 4, pp. 2563–2576, Aug. 2016.

[110] ——, "Pay or perish: The economics of premium peering," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 2, pp. 353–366, Feb. 2017.

[111] R. T. B. Ma, J. Wang, and D.-M. Chiu, "Paid prioritization and its impact on net neutrality," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 2, pp. 367–379, Feb. 2017.

[112] M. Zou, R. T. B. Ma, X. Wang, and Y. Xu, "On optimal service differentiation in congested network markets," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2693–2706, Dec. 2018.

[113] J. Tang and R. T. B. Ma, "Regulating monopolistic ISPs without neutrality," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 7, pp. 1666–1680, Jul. 2019.

[114] G. Davies, M. Hardt, and F. Kelly, "Come the revolution - network dimensioning, service costing and pricing in a packet switched environment," *Telecommunications Policy*, vol. 28, no. 5-6, pp. 391–412, Jun./Jul. 2004.

[115] J. Crowcroft, "Net neutrality: the technical side of the debate: a white paper," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 49–56, Jan. 2007.

[116] A. Meddeb, "Internet QoS: Pieces of the puzzle," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 86–94, Jan. 2010.

[117] N. Shetty, G. Schwartz, and J. Walrand, "Internet QoS and regulations," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 6, pp. 1725–1737, Dec. 2010.

[118] M. A. Callejo-rodriguez and J. Enriquez-Gabeiras, "Bridging the standardization gap to provide QoS in current NGN architectures," *IEEE Communications Magazine*, vol. 46, no. 10, pp. 132–137, Oct. 2008.

[119] R. Stankiewicz, P. Cholda, and A. Jajszczyk, "QoX: What is it really?" *IEEE Communications Magazine*, vol. 49, no. 4, Apr. 2011.

[120] F. Li, J. Sun, M. Zukerman, Z. Liu, Q. Xu, S. Chan, G. Chen, and K.-T. Ko, "A comparative simulation study of TCP/AQM systems for evaluating the potential of neuron-based AQM schemes," *Journal of Network and Computer Applications*, vol. 41, pp. 274–299, May 2014.

[121] Q. Xu, F. Li, J. Sun, and M. Zukerman, "A new TCP/AQM system analysis," *Journal of Network and Computer Applications*, vol. 57, pp. 43–60, Nov. 2015.

[122] Verizon, "Broadband service level agreement," Aug. 2017. [Online]. Available: http://www.verizonenterprise.com/external/service_guide/reg/cp_ibs_broadband_sla.pdf

[123] R. G. Addie, D. Fatseas, Y. Peng, F. Li, and M. Zukerman, "How good (or bad) is shortest path routing in layered networks," in *Proc. Australasian Telecommunication Networks and Applications Conference (ATNAC) 2012*, Brisbane, Queensland, Australia, Nov. 2012.

[124] Y. Peng, R. Lin, F. Li, C. Xing, J. Guo, W. Hu, V. Abramov, R. G. Addie, and M. Zukerman, "Validation of multi-layer network optimization," in *Proc. 2016 18th International Conference on Transparent Optical Networks (ICTON)*, Terento, Italy, Jul. 2016.

[125] C. Xing, R. G. Addie, Y. Peng, R. Lin, F. Li, W. Hu, V. M. Abramov, and M. Zukerman, "Resource provisioning for a multi-layered network," *IEEE Access*, vol. 7, pp. 16 226–16 245, 2019.

[126] M. Chryssomallis, "Smart antennas," *IEEE Antennas and Propagation magazine*, vol. 42, no. 3, pp. 129–136, Jun. 2000.

[127] O. J. Ciceri, C. A. Astudillo, and N. L. da Fonseca, "DBA algorithm with prioritized services for 10G-EPON with Multi-ONU customers," in *Proc. 2019 IEEE Latin-American Conference on Communications (LATINCOM)*, Bahia, Salvador, Nov. 2019.

[128] A. Veres, A. T. Campbell, M. Barry, and L.-H. Sun, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE Journal on selected Areas in Communications*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.

[129] B. E. Carpenter and K. Nichols, "Differentiated services in the Internet," *Proceedings of the IEEE*, vol. 90, no. 9, pp. 1479–1494, Sep. 2002.

[130] J. Postel, “Internet protocol,” RFC791, Sep. 1981. [Online]. Available: https://tools.ietf.org/html/rfc791

[131] P. Almquist, “Type of service in the internet protocol suite,” RFC1349, Jul. 1992. [Online]. Available: https://tools.ietf.org/html/rfc1349

[132] J. W. Forgie, “ST - a proposed Internet stream protocol,” ien119, Sep. 1979. [Online]. Available: https://www.rfc-editor.org/ien/ien119.txt

[133] C. Topolcic, “Experimental Internet stream protocol, version 2 (ST-II),” RFC1190, Oct. 1990. [Online]. Available: https://tools.ietf.org/html/rfc1190

[134] L. Delgrossi and L. Berger, “Internet stream protocol version 2 (ST2) protocol specification - version ST2+,” RFC1819, Aug. 1995. [Online]. Available: https://tools.ietf.org/html/rfc1819

[135] R. Braden, D. Clark, and S. Shenker, “Integrated services in the Internet architecture: an overview,” RFC1663, Jul. 1994. [Online]. Available: https://tools.ietf.org/html/rfc1663

[136] L. Zhang, S. Berson, S. Herzog, and S. Jamin, “Resource reservation protocol (RSVP) – version 1 functional specification,” RFC2205, Sep. 1997. [Online]. Available: https://tools.ietf.org/html/rfc2205

[137] J. Wroclawski, “The use of RSVP with IETF integrated services,” RFC2210, Sep. 1997. [Online]. Available: https://tools.ietf.org/html/rfc2210

[138] ——, “Specification of the controlled-load network element service,” RFC2211, Sep. 1997. [Online]. Available: https://tools.ietf.org/html/rfc2211

[139] S. Shenker, C. Partridge, and R. Guerin, “Specification of guaranteed quality of service,” RFC2212, Sep. 1997. [Online]. Available: https://tools.ietf.org/html/rfc2212

[140] F. Baker, C. Iturralde, F. L. Faucheur, and B. Davie, “Aggregation of RSVP for IPv4 and IPv6 reservations,” RFC3175, Sep. 2001. [Online]. Available: https://tools.ietf.org/html/rfc3175

[141] R. Wójcik, “Net neutral quality of service differentiation in flow-aware networks,” Ph.D. dissertation, AGH University of Science and Technology, 2011.

[142] R. Wójcik and A. Jajszczyk, “Flow oriented approaches to QoS assurance,” *ACM Computing Surveys*, vol. 44, no. 1, pp. 5:1–5:37, Jan. 2012.

[143] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An architecture for differentiated services,” RFC2475, Dec. 1998. [Online]. Available: https://tools.ietf.org/html/rfc2475

[144] K. Kilkki, *Differentiated services for the Internet*. Macmillan Publishing Co., Inc., 1999.

[145] S. G. Chaudhuri, “Design and implementation of a differentiated service based QoS model for real-time interactive traffic on constrained bandwidth IP networks,” Master’s thesis, Indian Institute of Technology, Kharagpur, Feb. 2010.

[146] F. L. Faucheur, L. Wu, B. Davie, S. Davari, P. Vaananen, R. Krishnan, P. Cheval, and J. Heinanen, “Multi-protocol label switching (MPLS) support of differentiated services,” RFC3270, May 2002. [Online]. Available: https://tools.ietf.org/html/rfc3270

[147] E. Osborne and A. Simha., “Traffic engineering with MPLS,” Cicso Press, Aug. 2002. [Online]. Available: https://www.ciscopress.com/articles/article.asp?p=28688

[148] *MPLS: Traffic Engineering: DiffServ Configuration Guide, Cisco IOS Release 15M&T*, Cisco Inc., Jan. 2018. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/mp_te_diffserv/configuration/15-mt/mp-te-diffserv-15-mt-book/mp-diffserv-tun-mode.html

[149] *Configuration Guide - MPLS S5700 and S6720 V200R013C00*, HUAWEI, Jun. 2020. [Online]. Available: https://support.huawei.com/enterprise/kz/doc/EDOC1100065729/369aa225/mpls-diffserv

[150] J. Crowcroft and P. Oechslin, “Differentiated end-to-end internet services using a weighted proportional fair sharing TCP,” *ACM SIGCOMM Computer Communication Review*, vol. 28, no. 3, pp. 53–69, Jul. 1998.

[151] N. Seddigh, B. Nandy, and P. Pieda, “Bandwidth assurance issues for TCP flows in a differentiated services network,” in *Proc. GLOBECOM’99*, vol. 3, Rio de Janeireo, Brazil, Dec. 1999, pp. 1792–1798.

[152] S. Sahu, P. Nain, C. Diot, V. Firoiu, and D. Towsley, “On achievable service differentiation with token bucket marking for TCP,” *ACM SIGMETRICS Performance Evaluation Review*, vol. 28, no. 1, pp. 23–33, Jun. 2000.

[153] A. Feroz, S. Kalyanaraman, and A. Rao, “A TCP-friendly traffic marker for IP differentiated services,” in *Proc. 2000 Eighth International Workshop on Quality of Service. IWQoS 2000 (Cat. No. 00EX400)*, Pittsburgh, PA, Jun. 2000, pp. 138–147.

[154] S.-H. Lee, S.-J. Seok, S.-J. Lee, and C.-H. Kang, “Study of TCP and UDP flows in a differentiated services network using two markers system,” in *Proc. 2001 IFIP/IEEE International Conference on Management of Multimedia Networks and Services*, Chicago, IL, Oct./Nov. 2001, pp. 198–203.

[155] Y. Yiakoumis, S. Katti, and N. McKeown, “Neutral net neutrality,” in *Proceedings of the 2016 ACM SIGCOMM Conference*, Florianópolis, Brazil, Aug. 2016, pp. 483–496.

[156] Huawei, “white-paper-for-the-traffic-blocking-theory-for-data-services-cn,” Mar. 2017. [Online]. Available: http://www-file.huawei.com/~/media/CORPORATE/PDF/whitepaper/white-paper-for-the-traffic-blocking-theory-for-data-services-cn.pdf

[157] A. Rao, A. Legout, Y.-s. Lim, D. Towsley, C. Barakat, and W. Dabbous, “Network characteristics of video streaming traffic,” in *Proceedings of the Seventh Conference on emerging Networking EXperiments and Technologies*, Tokyo, Japan, Dec. 2011.

[158] S. Sengupta, H. Gupta, P. De, B. Mitra, S. Chakraborty, and N. Ganguly, “Understanding data traffic behaviour for smartphone video and audio apps,” in *Proc. 2016 8th International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, Jan. 2016.

[159] H. Le Vu and M. Zukerman, "Blocking probability for priority classes in optical burst switching networks," *IEEE Communications Letters*, vol. 6, no. 5, pp. 214–216, May 2002.

[160] S. Yang and N. Stol, "Performance modeling in multi-service communications systems with preemptive scheduling," *Journal of Communications*, vol. 9, no. 6, pp. 448–460, Jun. 2014.

[161] L. Katzschner, "Loss sytstems with displacing priorities," in *Proceedings of 6th International Teletraffic Congress*, Munich, Germany, Sep. 1970.

[162] D. G. Kendall, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain," *The Annals of Mathematical Statistics*, pp. 338–354, 1953.

[163] M. Zukerman, "Introduction to queueing theory and stochastic teletraffic models." [Online]. Available: http://www.ee.cityu.edu.hk/~zukerman/classnotes.pdf

[164] C. Williamson, "Internet traffic measurement," *IEEE Internet computing*, vol. 5, no. 6, pp. 70–74, 2001.

[165] P. Loiseau, P. Gonçalves, G. Dewaele, P. Borgnat, P. Abry, and P. V.-B. Primet, "Investigating self-similarity and heavy-tailed distributions on a large-scale experimental facility," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1261–1274, 2010.

[166] S. Foss, D. Korshunov, S. Zachary *et al.*, *An introduction to heavy-tailed and subexponential distributions*. Springer, 2011, vol. 6.

[167] K. Sigman, "A primer on heavy-tailed distributions," *Queueing systems*, vol. 33, no. 1-3, p. 261, 1999.

[168] R. G. Addie, T. D. Neame, and M. Zukerman, "Performance evaluation of a queue fed by a Poisson Pareto burst process," *Computer Networks*, vol. 40, no. 3, pp. 377–397, Oct. 2002.

[169] ——, "Performance analysis of a Poisson-Pareto queue over the full range of system parameters," *Computer Networks*, vol. 53, no. 7, pp. 1099–1113, May 2009.

[170] J. Chen, R. G. Addie, M. Zukerman, and T. D. Neame, "Performance evaluation of a queue fed by a Poisson Lomax burst process," *IEEE Communications Letters*, vol. 19, no. 3, pp. 367–370, Mar. 2015.

[171] T. Neame, "Characterisation and modelling of internet traffic streams," Ph.D. dissertation, University of Melbourne, 2003.

[172] C. Campolo, A. Molinaro, A. Iera, and F. Menichella, "5G network slicing for vehicle-to-everything services," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 38–45, 2017.

[173] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[174] V. Sciancalepore, K. Samdanis, X. Costa-Perez, D. Bega, M. Gramaglia, and A. Banchs, "Mobile traffic forecasting for maximizing 5G network slicing resource utilization," in *Proc. INFOCOM 2017-IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[175] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.

[176] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[177] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5G and future mobile networks: mobility, resource management, and challenges," *IEEE communications magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[178] X. Li, M. Samaka, H. A. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, "Network slicing for 5G: Challenges and opportunities," *IEEE Internet Computing*, vol. 21, no. 5, pp. 20–27, 2017.

[179] S. Birmiwal, R. R. Mazumdar, and S. Sundaram, "Processor sharing and pricing implications," in *Proceedings of the 24th International Teletraffic Congress*, Kraków, Poland, Sep. 2012.

[180] F. M. Guillemin and R. R. Mazumdar, "Conditional sojourn times and the volatility of payment schemes for bandwidth sharing in packet networks," *Journal of Applied Probability*, vol. 52, no. 4, pp. 962–980, 2015.

[181] S. Birmiwal, R. R. Mazumdar, and S. Sundaram, "Pricing schemes in processor sharing systems," *Telecommunication Systems*, vol. 63, no. 3, pp. 421–435, Nov. 2016.

[182] "Network protocol analyzer – Wireshark." [Online]. Available: https://www.wireshark.org/

[183] R. Pauliks, I. Slaidins, K. Tretjaks, and A. Krauze, "Assessment of IP packet loss influence on perceptual quality of streaming video," in *Proc. 2015 Asia Pacific Conference on Multimedia and Broadcasting*, Bali, Indonesia, Apr. 2015.

[184] "Lillietest, statistics and machine learning toolbox," The Mathworks, Inc. [Online]. Available: https://www.mathworks.com/help/stats/lillietest.html

[185] *MATLAB version 9.8.0.1323502 (R2020a)*, The Mathworks, Inc., Natick, Massachusetts, 2020.

[186] "Worldwide broadband speed league 2018," 2018. [Online]. Available: https://www.cable.co.uk/broadband/speed/worldwide-speed-league/

[187] "Second-class netizens: outlying island residents get 100x slower internet speeds than rest of hong kong," Oct. 2015. [Online]. Available: https://www.scmp.com/news/hong-kong/education-community/article/1868808/island-residents-getting-left-behind-internet

[188] M. Cooper and B. Scott, “The importance of the internet and public support for network neutrality: national survey results,” Jan. 2006. [Online]. Available: https://consumerfed.org/pdfs/net_neutrality_poll.pdf

[189] J. K. Willcox, “Survey: Consumers favor strong net neutrality rules,” Sep. 2017. [Online]. Available: https://www.consumerreports.org/net-neutrality/most-consumers-still-want-strong-net-neutrality-rules/

[190] N. Scarborough, “Survey on net neutrality april 2018 questionnaire,” Apr. 2018. [Online]. Available: https://www.publicconsultation.org/wp-content/uploads/2018/04/Net_Neutrality_II_Quaire_041818.pdf

[191] R. Arnold, “The value of network neutrality to european consumers,” Apr. 2015. [Online]. Available: https://berec.europa.eu/eng/document_register/subject_matter/berec/download/2/5024-berec-report-on-how-consumers-value-net-_2.pdf

[192] A. Szabó and V. Pham, “Net neutrality and consumer demand in the video on-demand market,” May 2020. [Online]. Available: https://uh.edu/~aszabo2/cable8.pdf

[193] H. Aguinis and K. J. Bradley, “Best practice recommendations for designing and implementing experimental vignette methodology studies,” *Organizational Research Methods*, vol. 17, no. 4, pp. 351–371, Oct. 2014.

[194] O. James, S. R. Jilke, and G. G. Van Ryzin, *Experiments in public management research: Challenges and contributions*. Cambridge, United Kingdom: Cambridge University Press, 2017.

[195] G. F. Cavanagh and D. J. Fritzsche, “Using vignettes in business ethics research,” *Research in corporate social performance and policy*, vol. 7, pp. 279–293, 1985.

[196] R. M. Walker, M. Jin Lee, O. James, and S. M. Ho, “Analyzing the complexity of performance information use: Experiments with stakeholders to disaggregate dimensions of performance, data sources, and data types,” *Public Administration Review*, vol. 78, no. 6, Nov./Dec. 2018.

Appendix A

Graphs of Datasets

This parts contains all the graphs of datasets we used in the thesis. All the figures are drawn by Wireshark [182].

In all graphics, the unit (granularity) of the horizontal axes is milliseconds, and the unit of the vertical axes is bits. Green traffic represents TCP packets, magenta traffic represents UDP packets, and blue traffic represents ICMP packets.

A.1 WIDE project (Japan) datasets

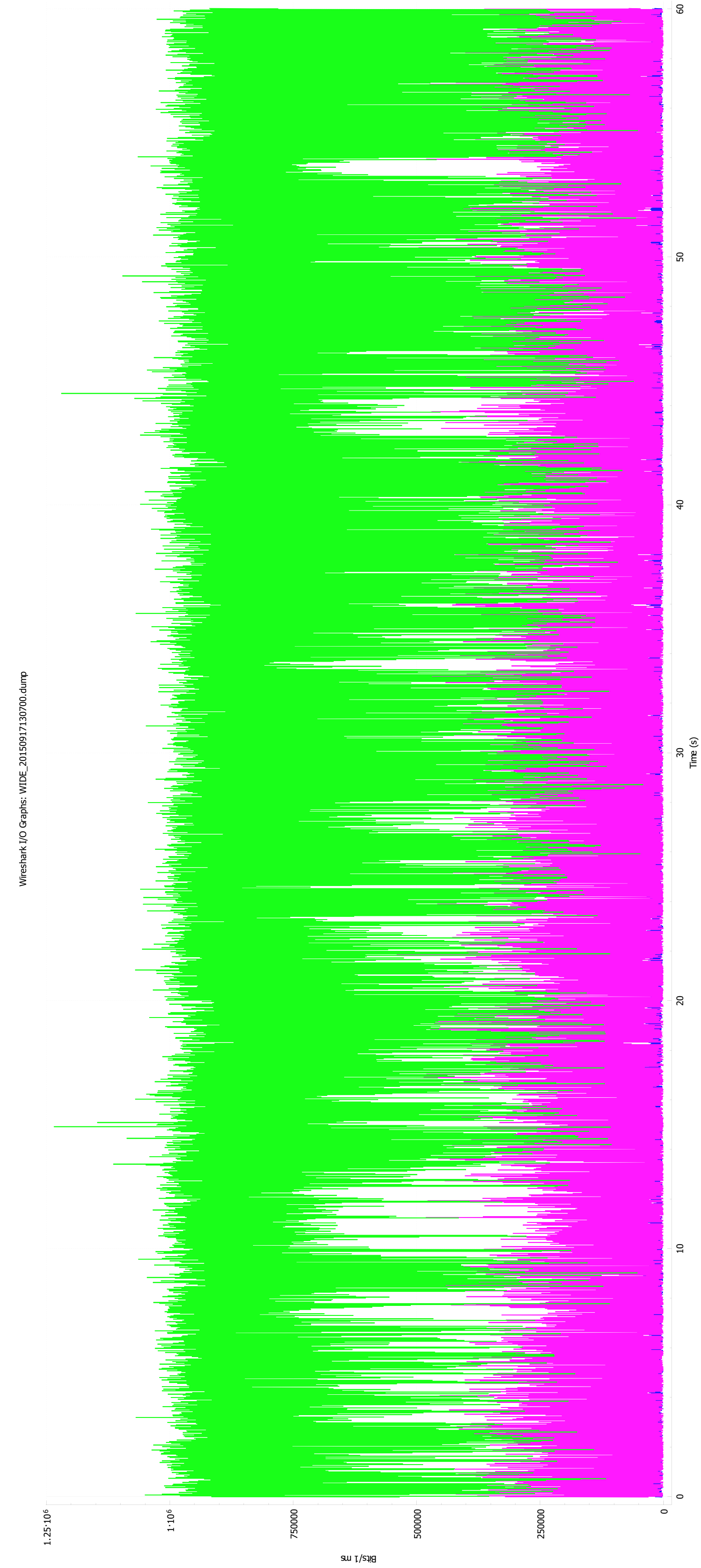


Fig. A.1 WIDE Sample F 20150917

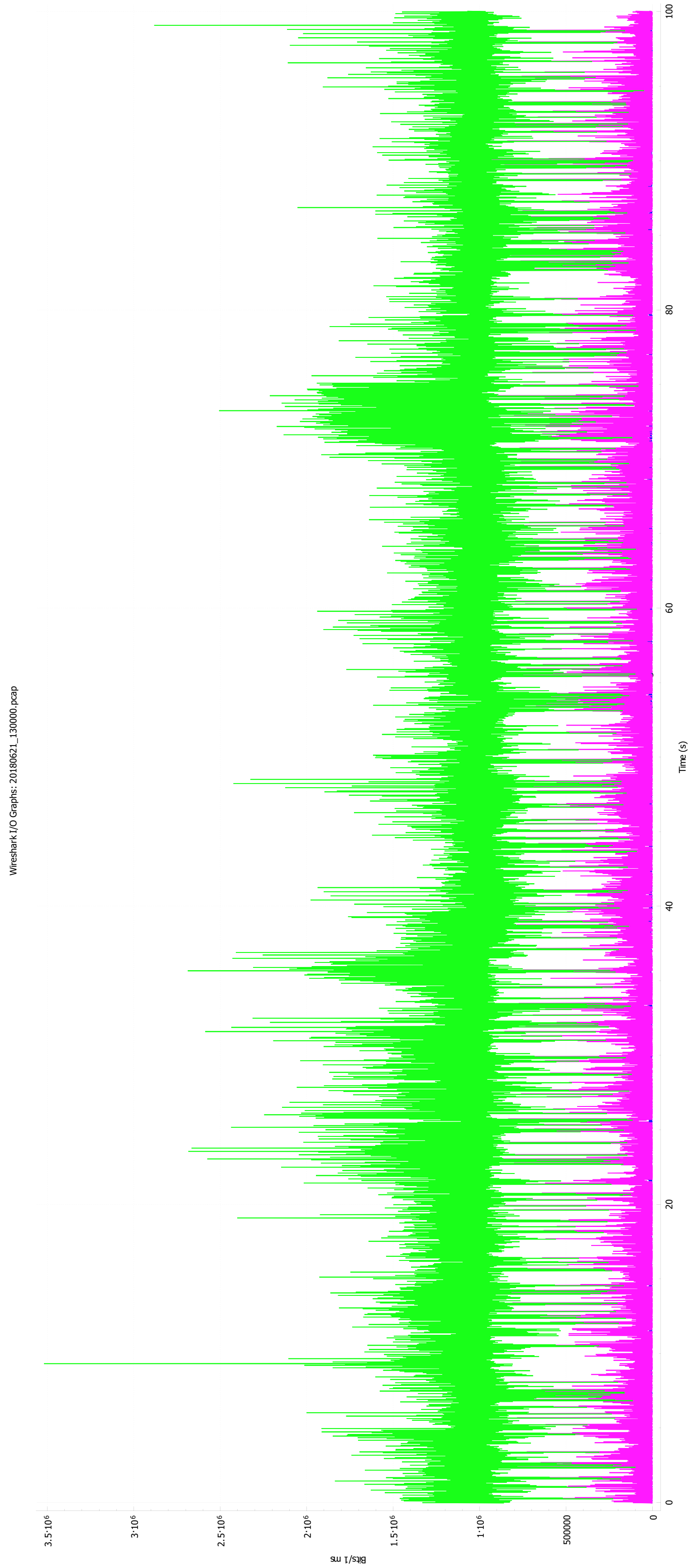


Fig. A.2 WIDE Sample F 20180621

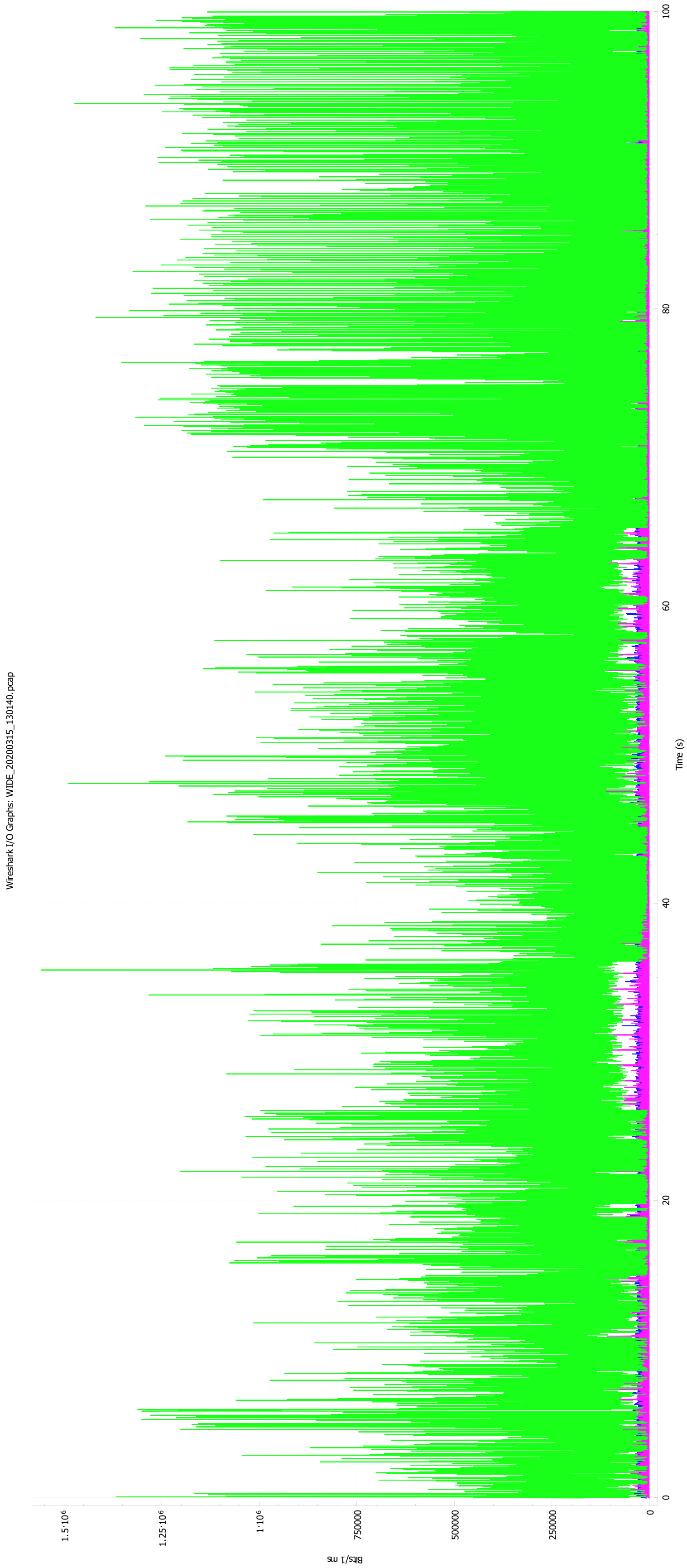


Fig. A.3 WIDE Sample F 20200315

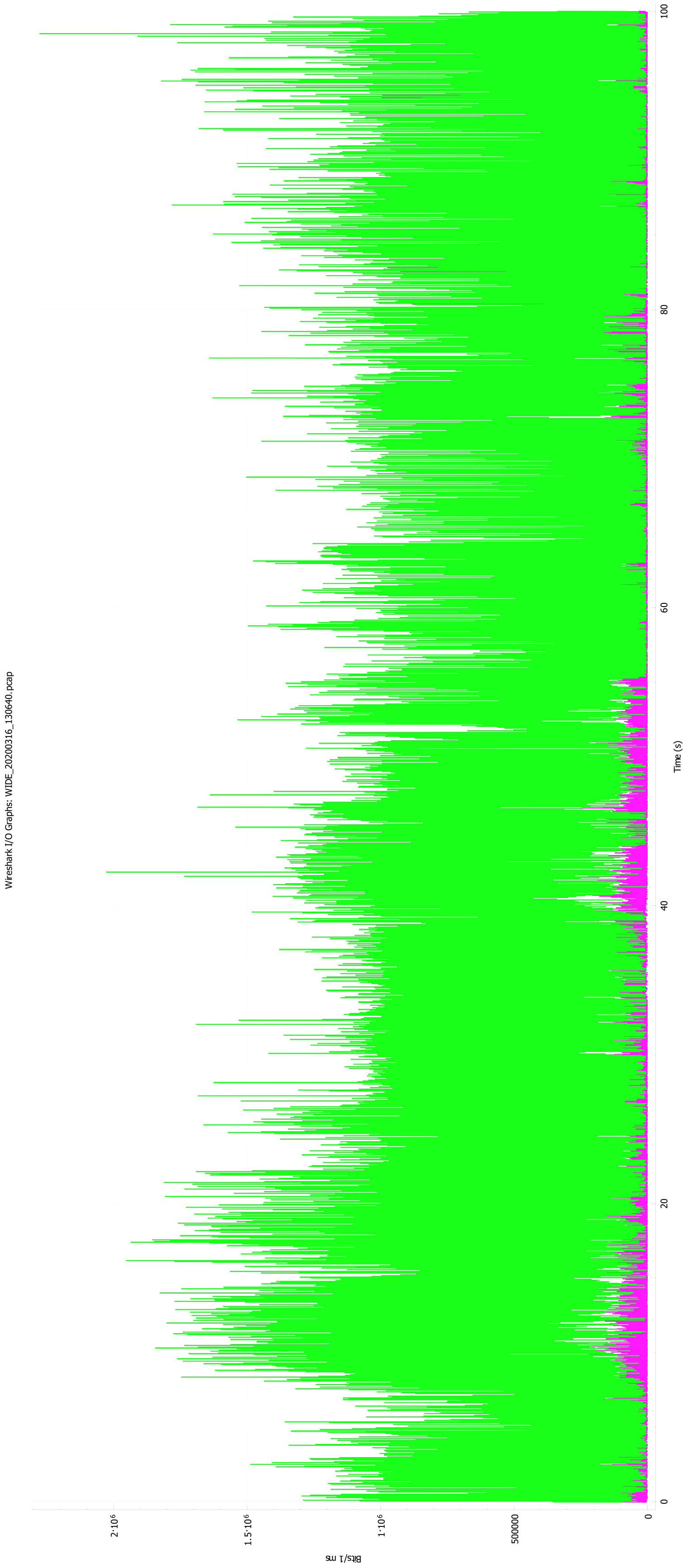


Fig. A.4 WIDE Sample F 20200316 A

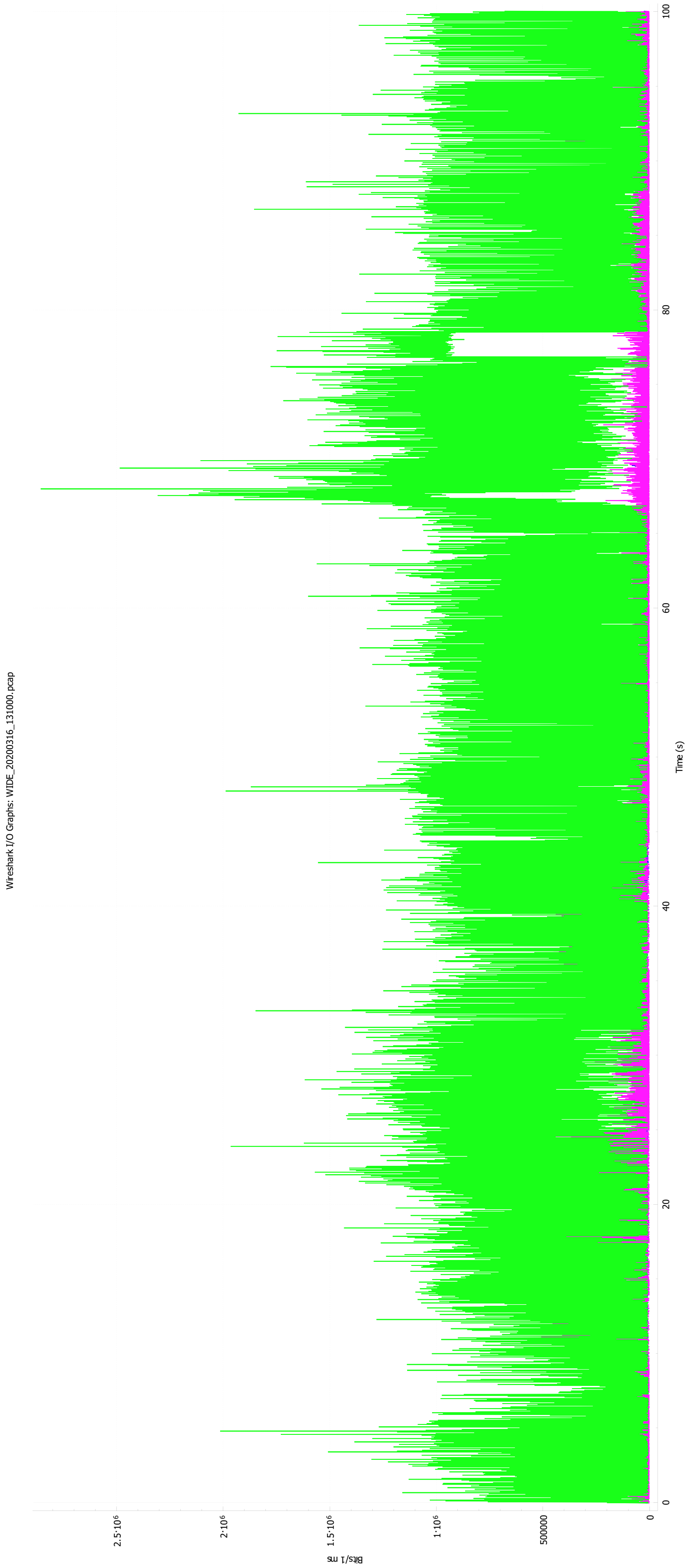


Fig. A.5 WIDE Sample F 20200316 B

A.2 Chicago City datasets 2015

Fig. A.6 Equinix-Chicago 20150219

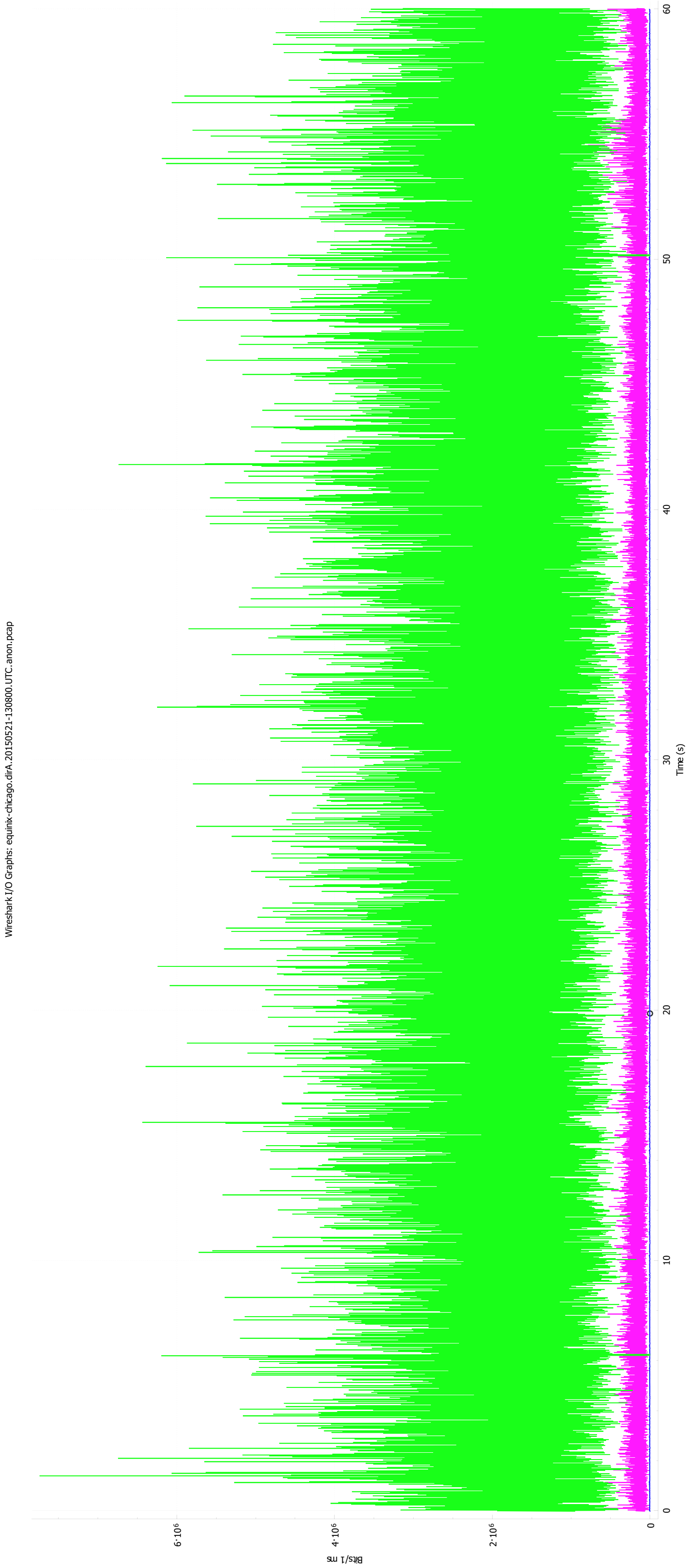


Fig. A.7 Equinix-Chicago 20150521

Fig. A.8 Equinix-Chicago 20150917

Fig. A.9 Equinix-Chicago 20151217

A.3 Chicago City datasets 2016

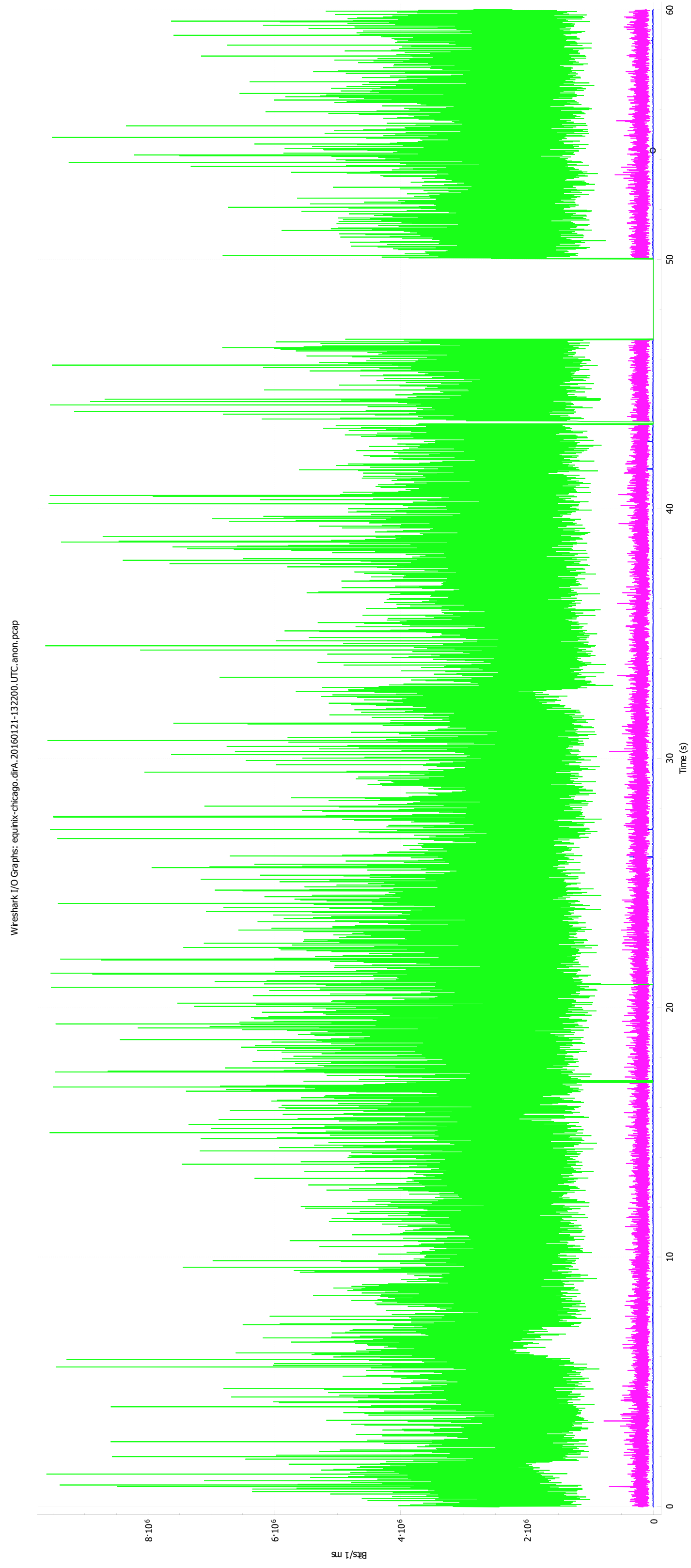


Fig. A.10 Equinix-Chicago 20160121

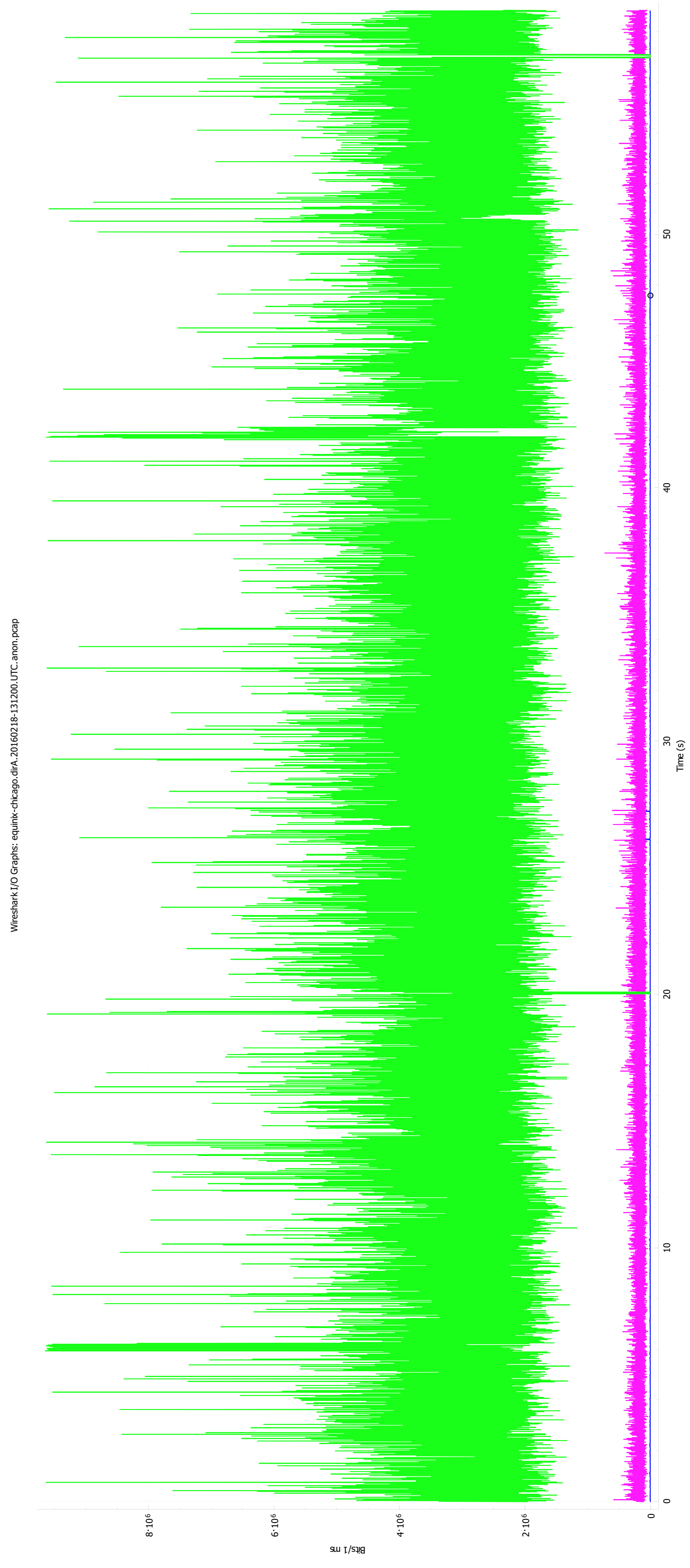


Fig. A.11 Equinix-Chicago 20160218

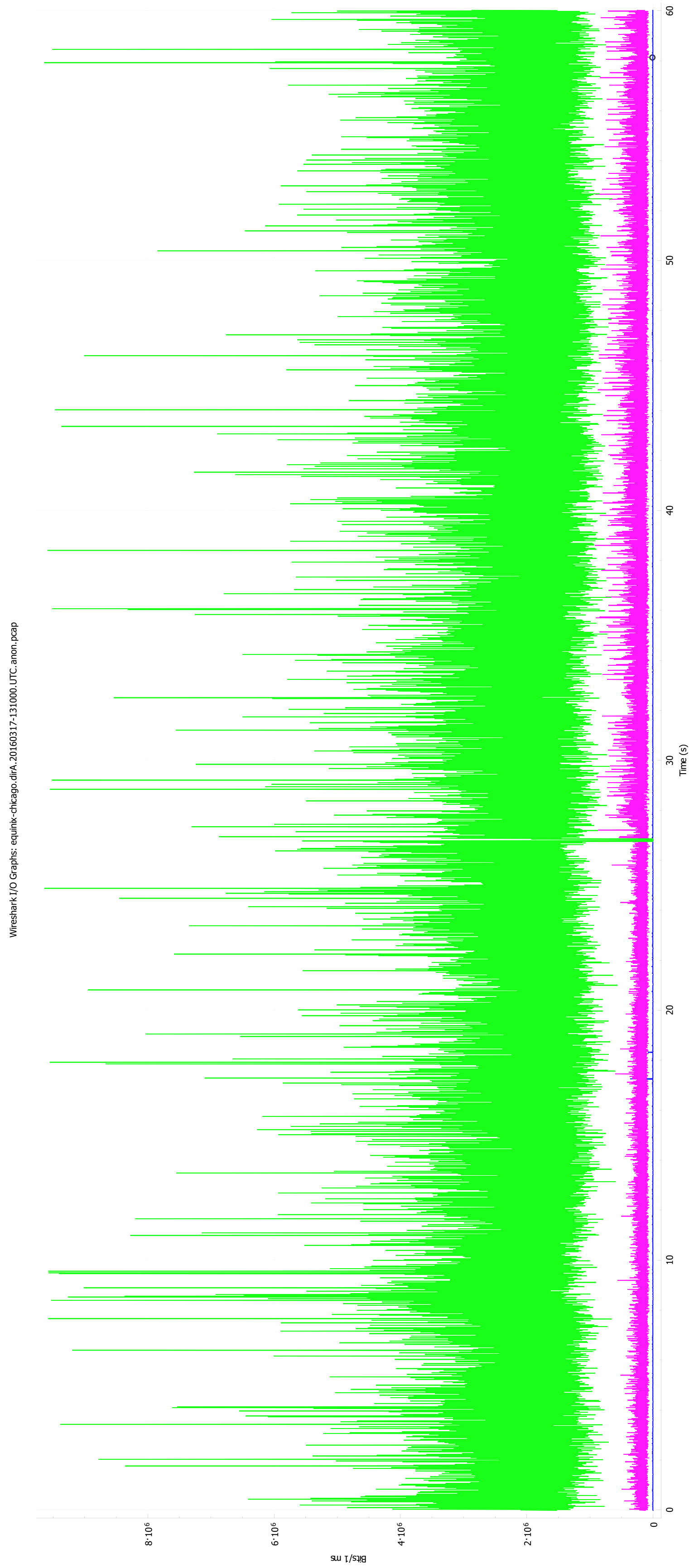


Fig. A.12 Equinix-Chicago 20160317

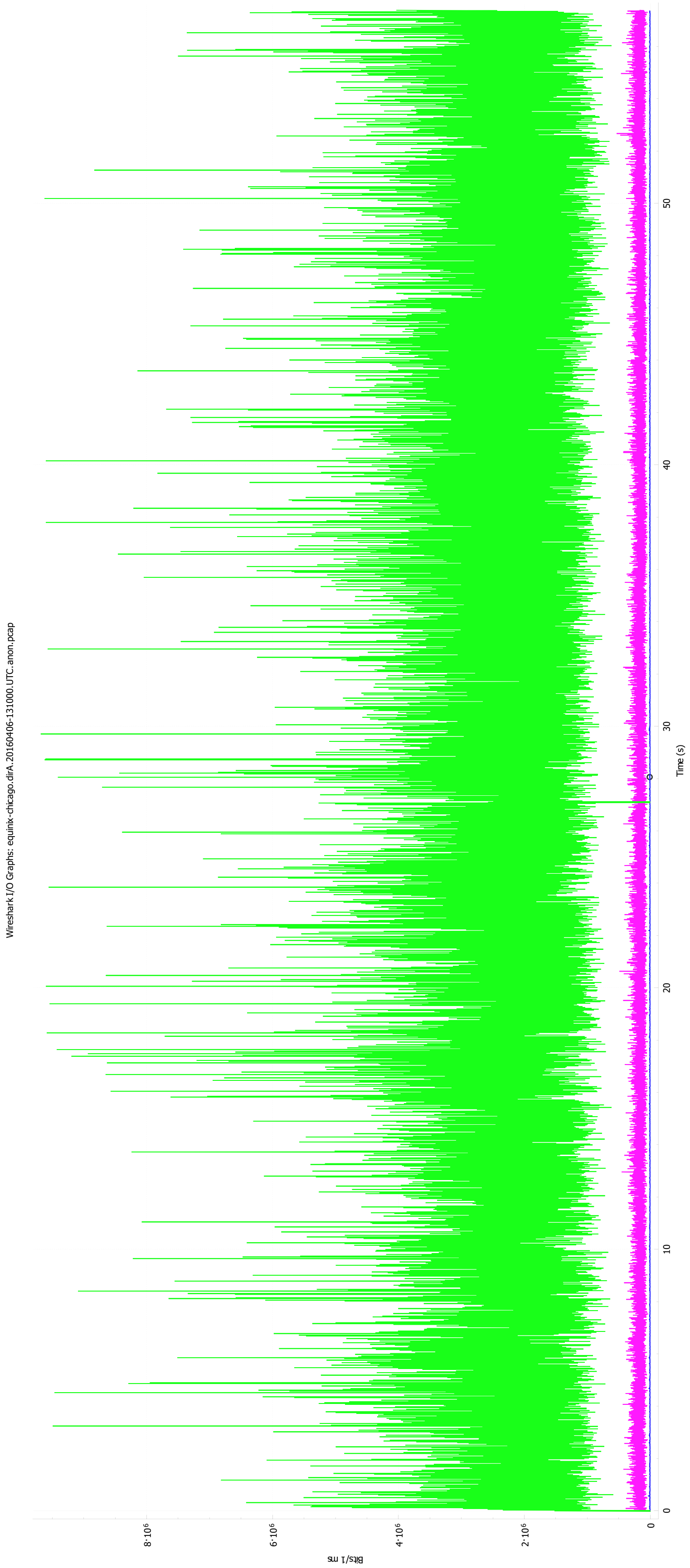


Fig. A.13 Equinix-Chicago 20160406

A.4 New York City datasets 2018

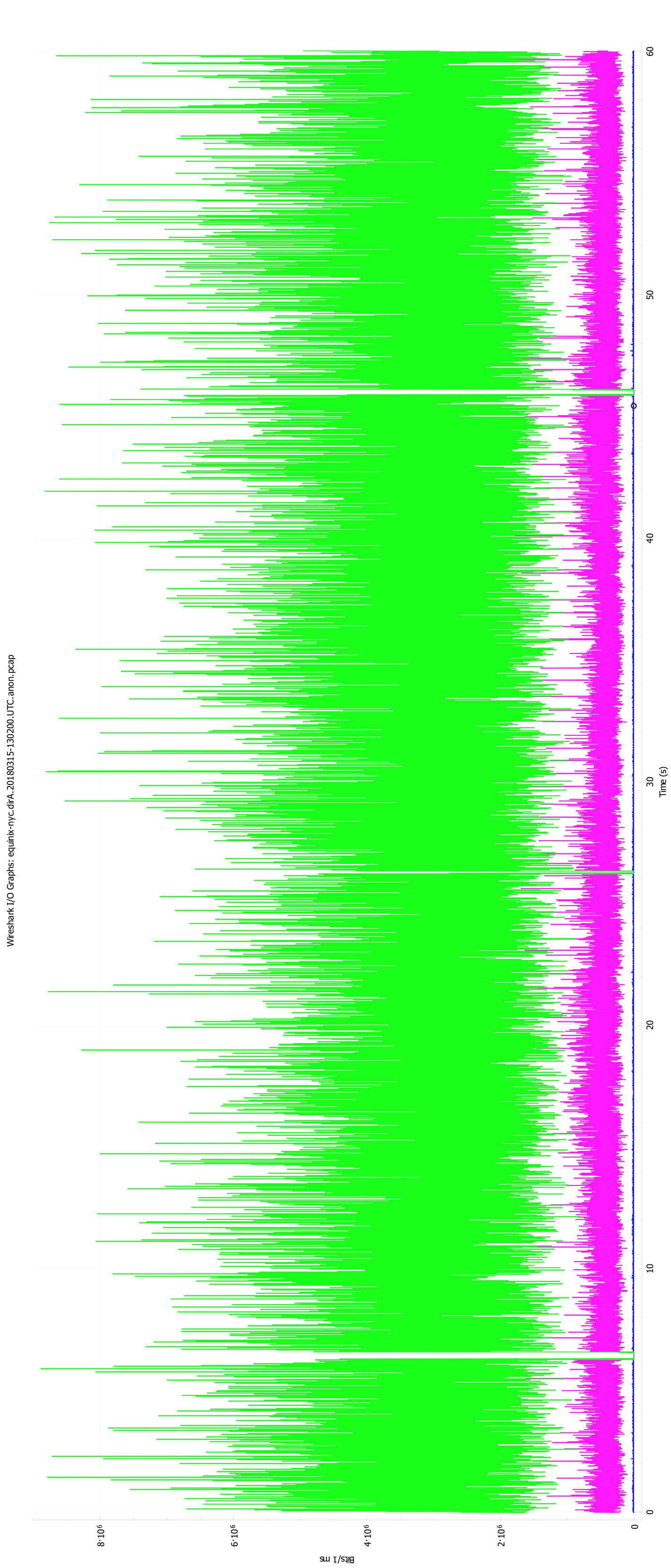


Fig. A.14 Equinix-New York 20180315

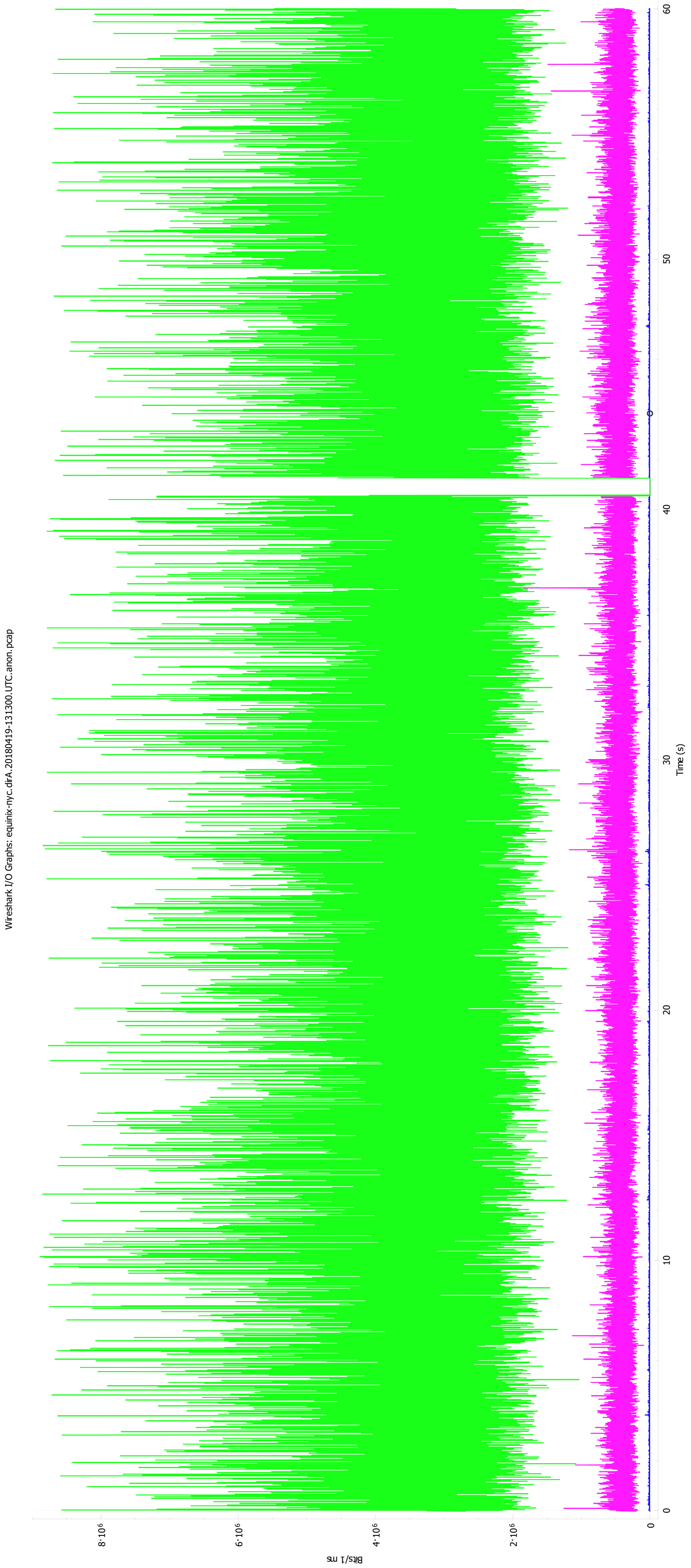


Fig. A.15 Equinix-New York 20180419

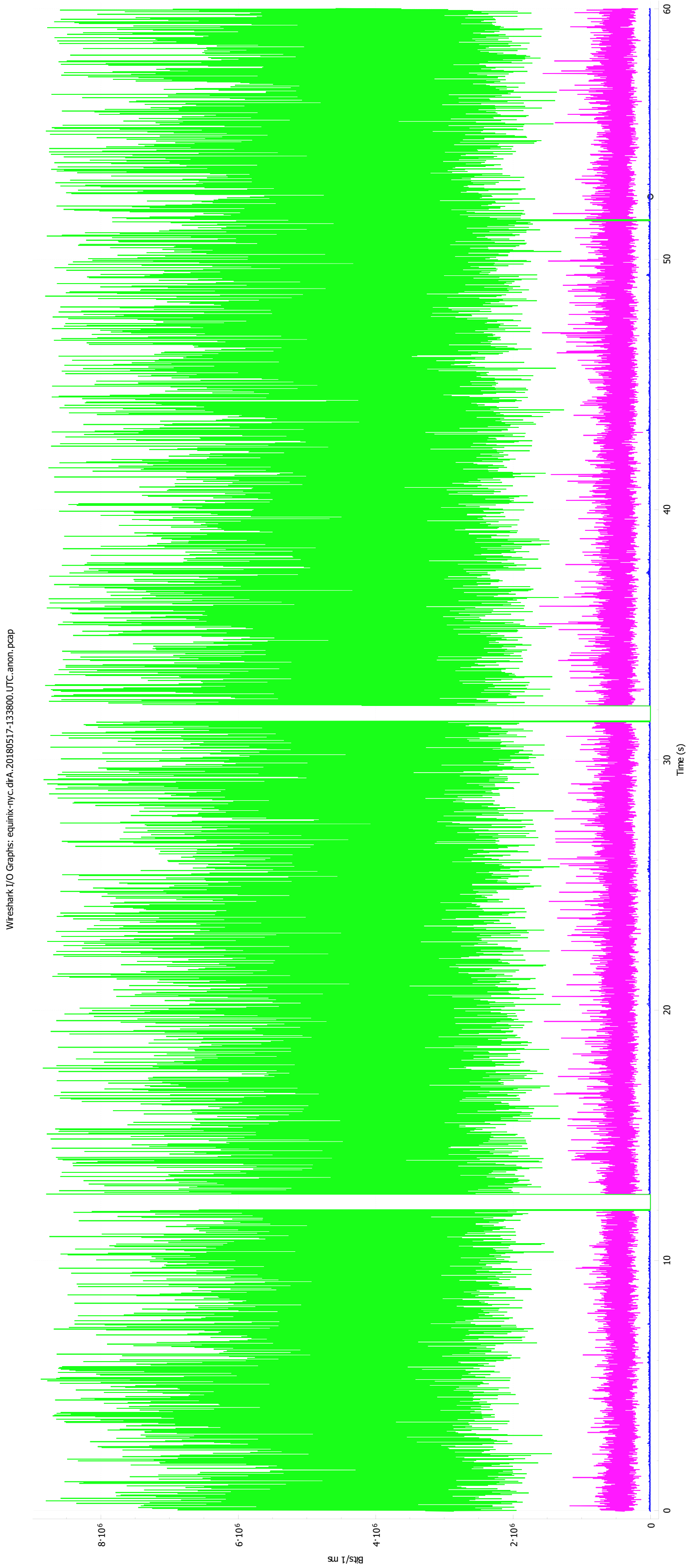


Fig. A.16 Equinix-New York 20180517

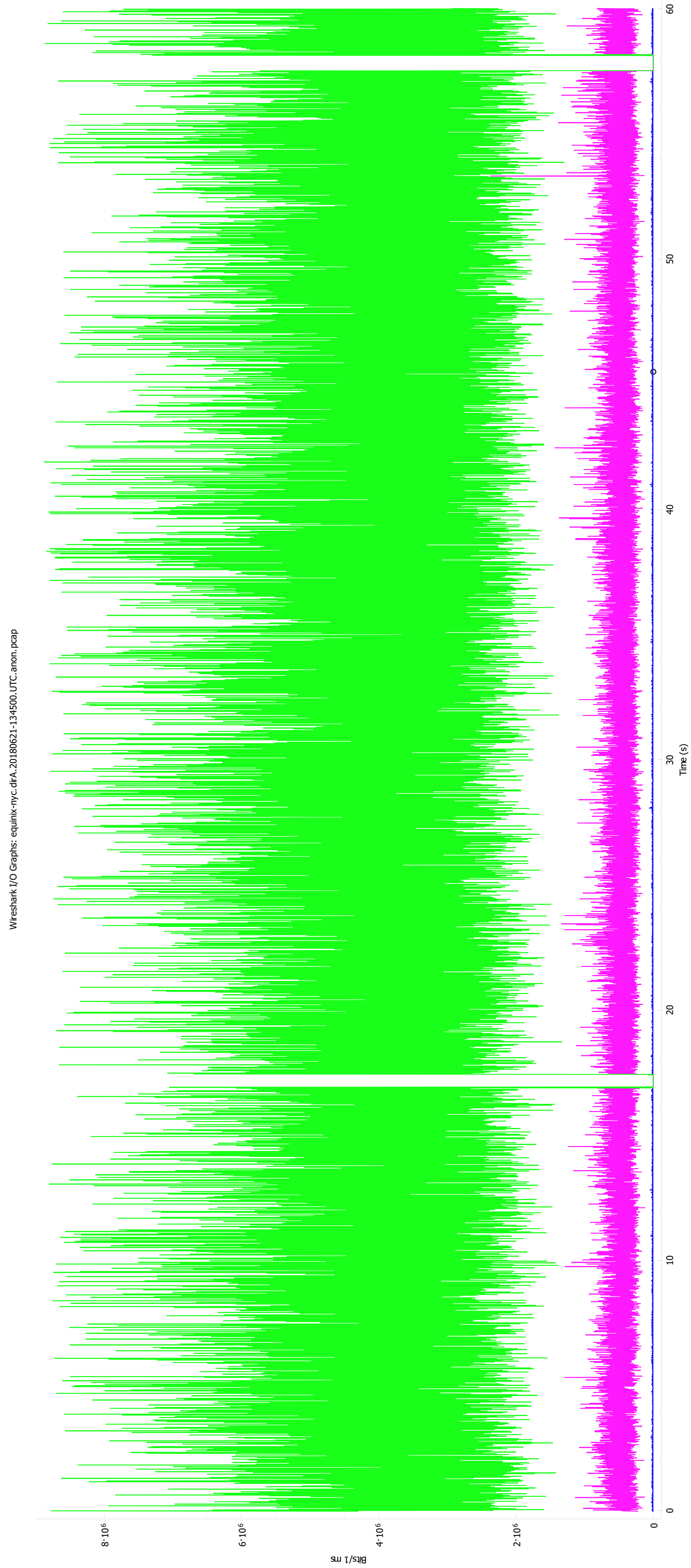


Fig. A.17 Equinix-New York 20180621

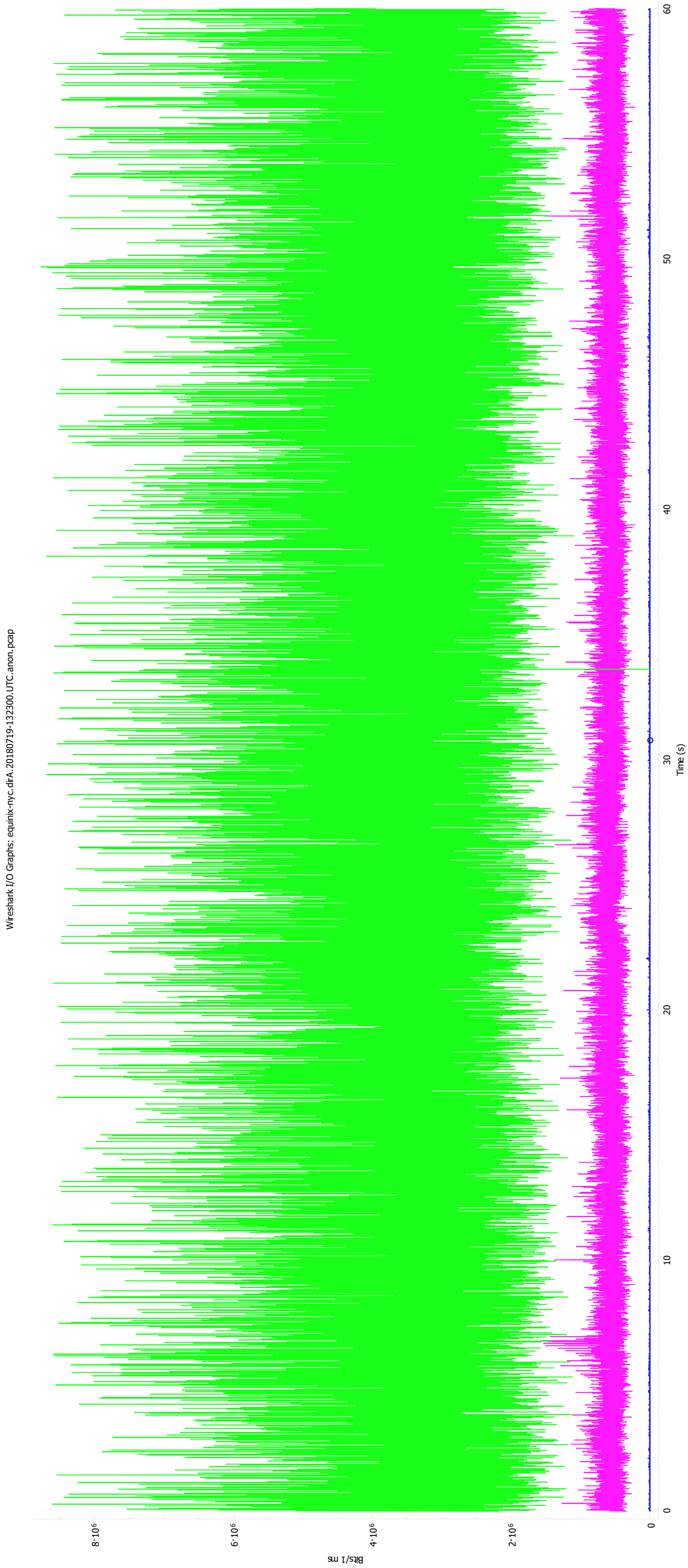


Fig. A.18 Equinix-New York 20180719

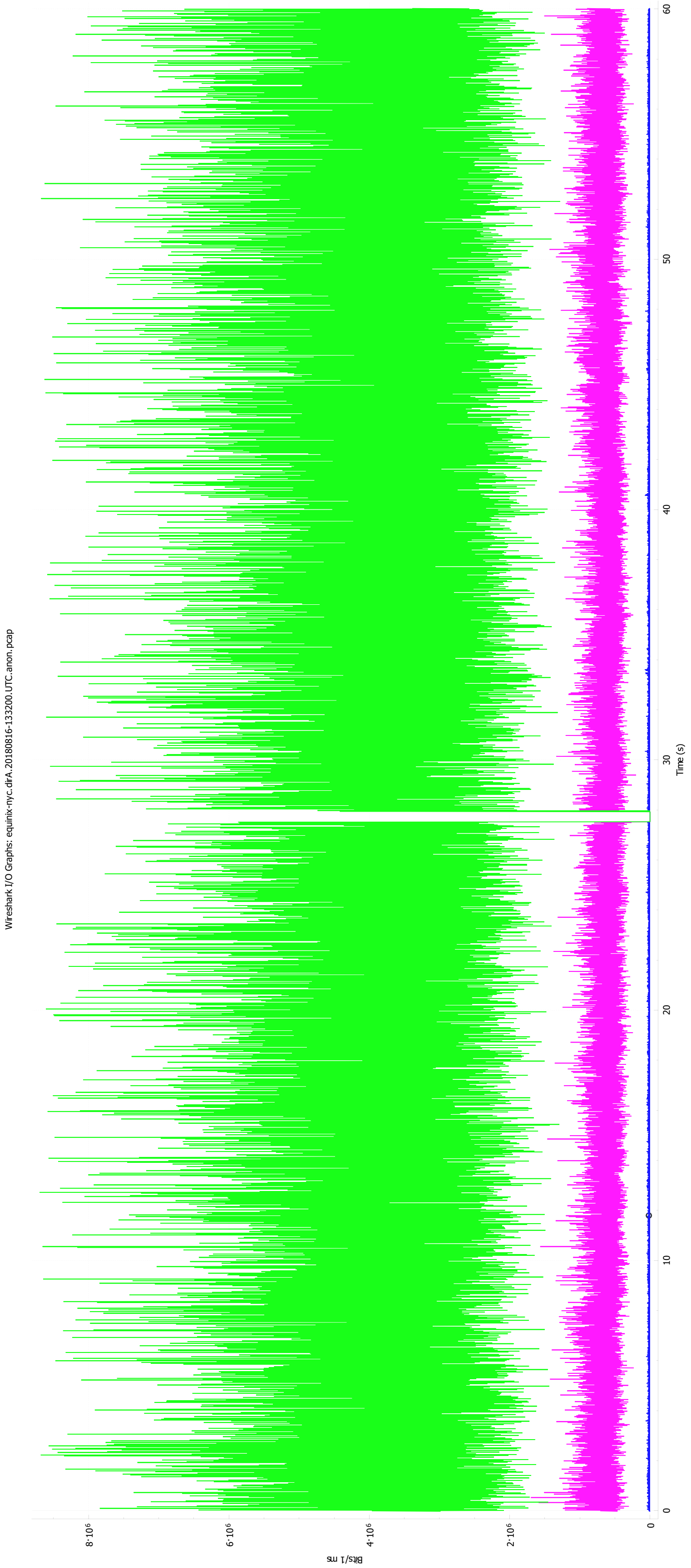


Fig. A.19 Equinix-New York 20180816

Appendix B

A Public Survey of Network Regulation Policy in Hong Kong

B.1 Terminologies and related information:

Compare to other public issues such as “Food safety” and “Air pollution”, the terminology concerning telecommunication network and concepts of network regulation policy are elusive and relatively complex for the public to understand. Therefore the following terminologies and related information are presented for helping.

In the next 4 pages respondents of the survey will be provided with a number of definitions of ideas and technologies associated with the Internet. Please read these carefully, and then answer the questions that follow.

Internet service provider: An organization that provides services for accessing, using, or participating in the internet. (Example: PCCW, 3 HK, CSL, Hong Kong Broadband, etc.)

Content provider: An organization that provides information for use on a website or in a mobile phone App (Example: Facebook, YouTube, Netflix).

Bandwidth hungry services: Applications that require large amount of data transmission and therefore consume significant amount of network resources (See Fig. B.1 below for example).

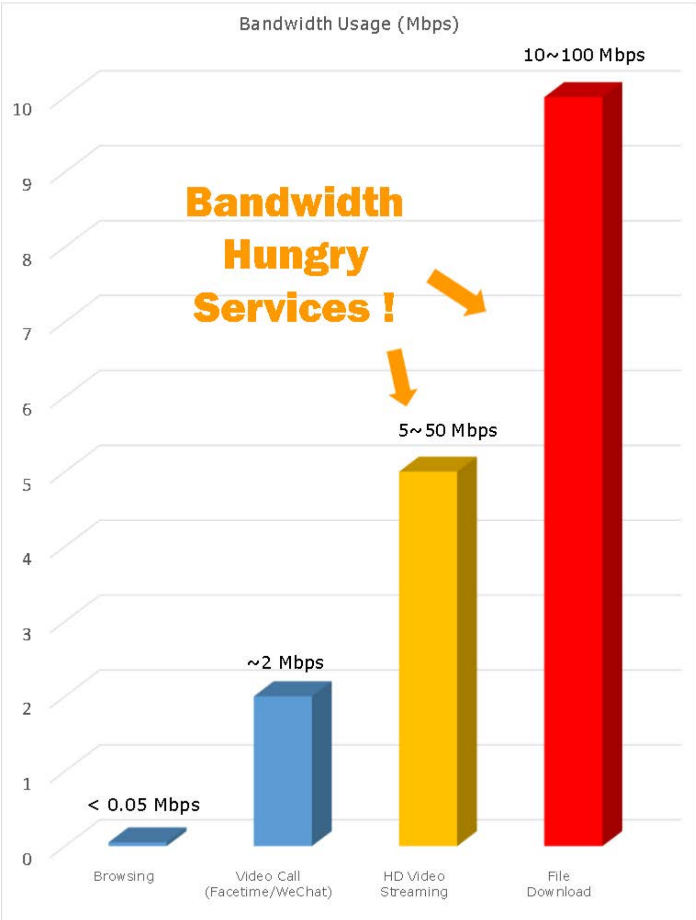


Fig. B.1 Example of bandwidth hungry service

Zero-rating: A benefit that an internet service provider may offer to their subscribers, who are able to access certain websites, services or applications without being charged, also called “toll-free” (Example: The Wikipedia website is free to access when using mobile devices in some countries, see Fig. B.2, just for illustration, not the case in Hong Kong).

Network congestion: A decrease in quality of service due to customers’ data demand exceeding network capability (Example: too many customers using the mobile network

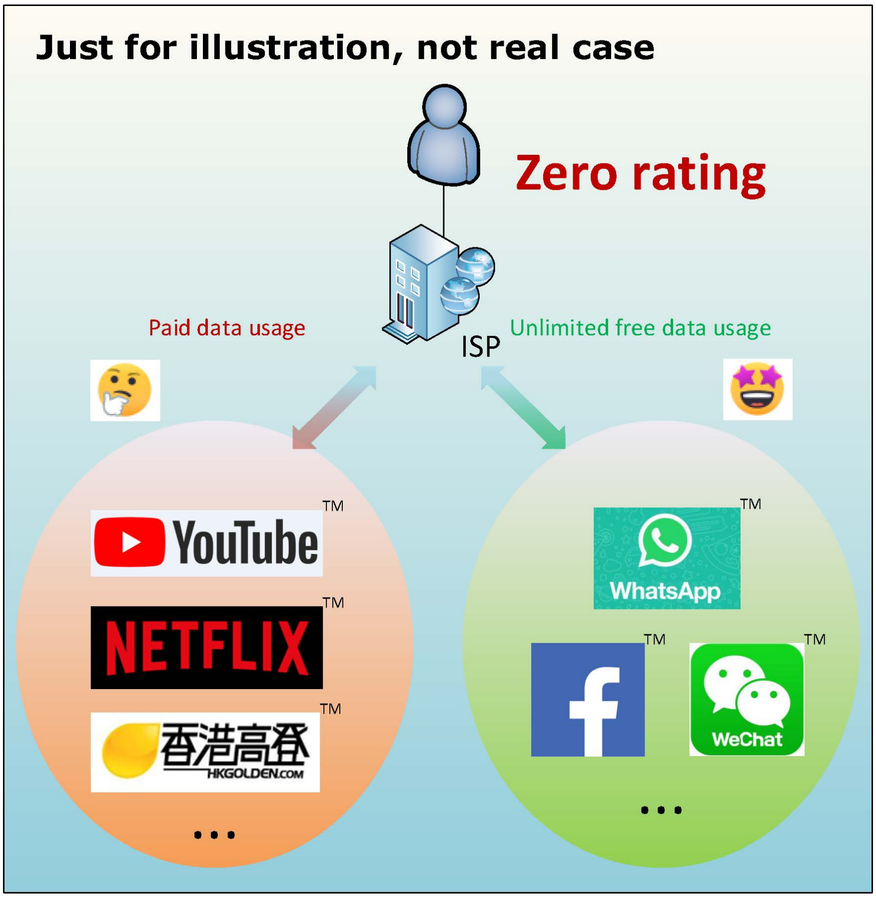


Fig. B.2 Illustration of zero-rating

during rush hour). Customers may suffer longer delays, loss of data, or even denial of access to websites or Apps.

Bandwidth throttling: An action by an internet service provider to increase or decrease the speed of an internet service.

Bandwidth throttling under and not under network congestion : Bandwidth throttling can be done either under congestion or no congestion conditions. Under congestion, the internet service provider may allow customers who pay more to have priority over other customers, so the traffic of those customers who pay less will suffer a longer delay or even lose connection. Under no congestion, internet service providers may reduce the rate of certain customers, who for example exceed the monthly data cap according to their plan, which can be higher or lower depending on the payment. See Fig. B.3 for example.

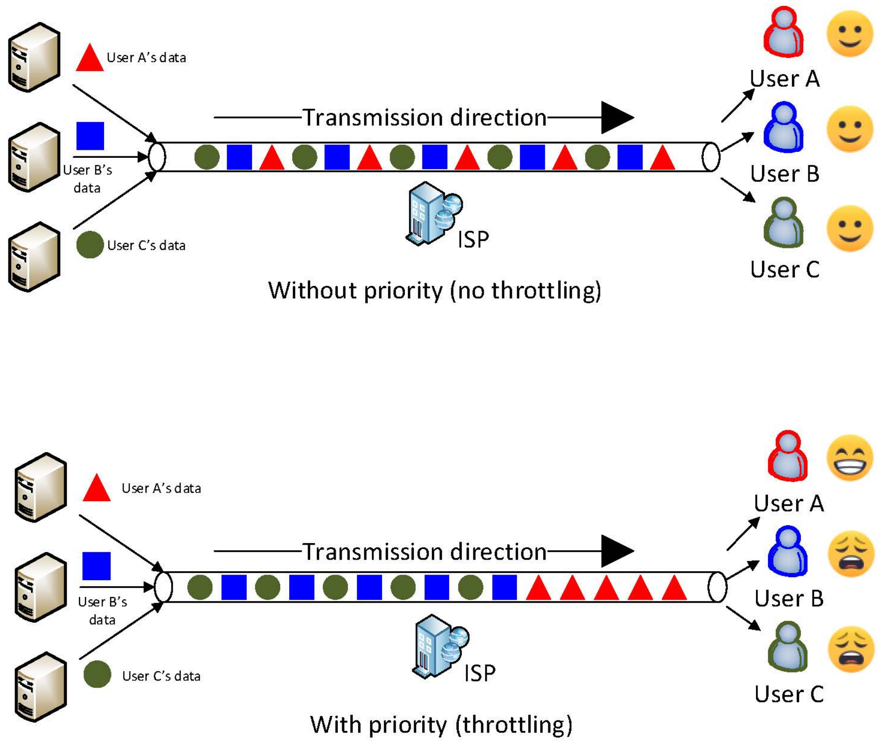


Fig. B.3 ISP throttling between different users

The need for investments: Increased internet traffic demands require ongoing investment in infrastructure upgrades (e.g., laying new optical fibre, upgrading and increasing the number of mobile base stations and upgrading switches and routers).

Network resource allocation: How to allocate the network resources fairly and efficiently is an important issue of public policy.

Justification for service prioritization: If internet service providers are allowed to charge more for better service (to give priority to premium customers), they will have the incentive and the resources to upgrade their network infrastructures.

Net neutrality: Proponents of net neutrality believe that the same treatment should apply to the delivery of all internet services to all users irrespective of their origin, content and destination. Examples of violation of net neutrality include: internet service providers providing the services of certain content providers (such as social media application) for free (so-called zero-rating), or give priority to those customers who pay more but limit/delay the customer who pay less (so-called throttling).

B.2 Preliminary questions

Before asking the questions of the network regulation policy, some preliminary questions are used to collect some general information on the habits of respondents. The preliminary questions are listed below:

1. Network Type: By which network do you reach internet?

- Only wireless (mobile network) (1)
- Only wireline (Wi-Fi is considered as wireline network) (2)
- Both Wireline and Wireless (3)
- Others (Please specify:______) (4)

2. Expenditure_1: How much do you pay for your wireline (Wi-Fi) network each month?

- None (1)
- 0 200 HKD (2)
- 200 500 HKD (3)
- Over 500 HKD (4)

3. Expenditure_2 How much do you pay for your wireless (mobile) network each month?

- None (1)
- 0 200 HKD (2)
- 200 500 HKD (3)
- Over 500 HKD (4)

4. Usage level: On average, how long do you spend on internet each day?

- None (1)
- 0 1 Hour (2)
- 1 4 Hours (3)
- 4 8 Hours (4)
- Over 8 Hours (5)

5. Net_Satisfy: How satisfied are you with your internet service provider (e.g. PCCW, 3 HK, CSL, Hong Kong Broadband, etc.)?

- Extremely dissatisfied (1)
- Moderately dissatisfied (2)
- Slightly dissatisfied (3)
- Neither satisfied nor dissatisfied (4)
- Slightly satisfied (5)
- Moderately satisfied (6)
- Extremely satisfied (7)

B.3 Four vignettes

In the following, four different vignettes are given which describe the setting strategy of the broadband market for Hong Kong. Respondents of the questionnaire receive one of these four vignettes randomly:

1. Under strict net neutrality regulation, ISPs are not allowed to provide any zero-rating services. This will prevent some big content providers using their market power to eliminate competition, while maintaining a free market environment. Meanwhile, internet service providers are not allowed to give priority to customers who pay more and throttle the rate of customers who pay less. In this case, maintaining a fair network service, the cost of government regulations and introducing complaints scheme are necessary. On the other hand, not allowing prioritization of services through bandwidth throttling will decrease the incentive of internet service providers to upgrade the telecommunication infrastructures to meet the ever-growing communication demand.

2. An internet service provider may introduce zero-rating services (free access) to certain selected social media services (e.g., WeChat, Facebook, WhatsApp). However, internet service providers are not allowed to prioritize in terms of access rates (through bandwidth throttling) between customers under network congestion, so that all customers may suffer longer delay during congestion period. In this case, customers may enjoy free-riding for some social media services. Potentially, the big content provider companies may use their power to eliminate competition, which may violate competition laws. Since no prioritized service through bandwidth throttling is allowed, the incentive of internet service providers to upgrade their networks may be low.

3. The internet service providers are not allowed to introduce zero-rating services (free access) to certain selected social media services (e.g., WeChat, Facebook, WhatsApp). However, they are allowed to prioritize in terms of access rates (through bandwidth

throttling) between customers under network congestion. Specifically, internet service providers may give priority to customers who pay more and reduce the access rate of the customers who pay less. Therefore, premium customers will enjoy a better experience while other customers may suffer longer delays. In this case, big content providers cannot use their market power to eliminate the competition. The use of prioritized service through bandwidth throttling will provide additional income and incentives to internet service providers to upgrade their infrastructures. The customers who pay more will enjoy the better internet service.

4. Internet service providers are allowed to introduce zero-rating services (free access) to certain selected social media services (e.g., WeChat, Facebook, WhatsApp). They are also allowed to prioritize in terms of access rates (through bandwidth throttling) between customers under network congestion. Specifically, internet service providers may give priority to customers who pay more and reduce the access rate of customers who pay less. Therefore, premium customers will enjoy a better experience while other customers may suffer longer delays. In this case, big content providers may use their market power to eliminate the competition, and internet service providers have a high incentive to upgrade their network infrastructures.

After each vignette being sent to the respondent, the following 6 questions are used to collect their opinions:

Vignette Question 1: I am satisfied with my current access rate to the internet.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)

- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Vignette Question 2: To what extent do you agree that internet service providers should provide zero-rating service?

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Vignette Question 3: To what extent do you agree that this is an effective strategy to maintain the free-market place of Hong Kong?

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)

- Strongly agree (7)

Vignette Question 4: To what extent do you agree that internet service providers should provide a prioritized service through bandwidth throttling so that customers who pay more get a better access rate to the internet?

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Vignette Question5: To what extent do you agree that this is an effective strategy for internet service providers to invest and upgrade their network infrastructures?

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

Vignette Question 6: To what extent do you agree that net neutrality is important to Hong Kong?

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

B.4 Net neutrality knowledgeability

The last question of this survey is to test the respondent's knowledge about the net neutrality policy. The definition of "zero-rating" is selected as the question.

Zero rating is defined as:

- Increased internet traffic demands require ongoing investment in infrastructure upgrades (e.g., laying new optical fibre, upgrading and increasing the number of mobile base stations and upgrading switches and routers). (1)
- A benefit that an internet service provider may offer to their subscribers, who are able to access certain websites, services or applications without being charged. (2)
- An action by an internet service provider to increase or decrease the speed of an internet service. (3)

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q1	124	530	4.2742	2.1193
V2_Q1	128	553	4.3203	2.2037
V3_Q1	125	579	4.6320	1.7183
V4_Q1	120	546	4.5500	2.1824

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11.3	3	3.7664	1.8324	0.1403	2.6230
Within Groups	1013	493	2.0554			
Total	1025	496				

Question 1: I am satisfied with my current access rate to the internet

Table B.1 The ANOVA analysis of access rate satisfaction

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q2	124	559	4.5081	1.9593
V2_Q2	128	618	4.8281	2.1120
V3_Q2	125	605	4.8400	2.1032
V4_Q2	120	558	4.6500	1.9101

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	9.385	3	3.1285	1.5465	0.2016	2.6230
Within Groups	997.3	493	2.0229			
Total	1007	496				

Question 2: To what extent do you agree that internet service providers should provide zero-rating service?

Table B.2 The ANOVA analysis of Zero-rating service agreement

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q3	124	581	4.6855	1.4694
V2_Q3	128	598	4.6719	1.6395
V3_Q3	125	618	4.9440	1.5210
V4_Q3	120	555	4.6250	1.7153

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.723	3	2.5743	1.6236	0.1830	2.6230
Within Groups	781.7	493	1.5856			
Total	789.4	496				

Question 3: To what extent do you agree that this is an effective strategy to maintain the free-market place of Hong Kong?

Table B.3 The ANOVA analysis of Free-market strategy agreement

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q4	124	510	4.1129	2.2961
V2_Q4	128	537	4.1953	2.5521
V3_Q4	125	531	4.2480	1.8493
V4_Q4	120	529	4.4083	2.0924

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5.667	3	1.8889	0.8584	0.4626	2.6230
Within Groups	1085	493	2.2005			
Total	1091	496				

Question 4: To what extent do you agree that internet service providers should provide a prioritized service through bandwidth throttling so that customers who pay more get a better access rate to the internet?

Table B.4 The ANOVA analysis of prioritized service agreement

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q5	124	575	4.6371	1.4851
V2_Q5	128	588	4.5938	1.8337
V3_Q5	125	577	4.6160	1.5288
V4_Q5	119	548	4.6050	1.6308

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.128	3	0.0427	0.0263	0.9942	2.6230
Within Groups	797.5	492	1.6210			
Total	797.7	495				

Question 5: To what extent do you agree that this is an effective strategy for internet service providers to invest and upgrade their network infrastructures?

Table B.5 The ANOVA analysis of ISP incentive strategy agreement

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
V1_Q6	124	636	5.1290	1.1377
V2_Q6	128	667	5.2109	1.4748
V3_Q6	125	655	5.2400	1.2968
V4_Q6	120	635	5.2917	1.2167

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.6993	3	0.5664	0.4413	0.7236	2.6230
Within Groups	632.8318	493	1.2836			
Total	634.5312	496				

Question 6: To what extent do you agree that net neutrality is important to Hong Kong?

Table B.6 The ANOVA analysis of net neutrality policy agreement