ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MANUSCRIPT-BASED THESIS PRESENTED TO
ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

BY
Ahmad Quttoum

RESOURCE MANAGEMENT FOR VIRTUALIZED NETWORKS

MONTREAL, OCTOBER 7 2011

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS:

Mr. Zbigniew Dziong, Thesis director
Département de génie électrique à l' École de technologie supérieure

Mr. Hadi Otrok, Thesis co-director (external)
Department of computer engineering at Khalifa university of science, technology & research

Mr. Witold Suryn, Committee president
Département de génie logiciel et des T.I. à l' École de technologie supérieure

Mrs. Anjali Agarwal, External examiner
Department of electrical and computer engineering at Concordia university

Mr. Michel Kadoch, Examiner
Département de génie électrique à l' École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

ON OCTOBER 5 2011

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

No matter what I say, I can't express my deep and sincere gratitude to Almighty ALLAH. Had it not been for his endless blessings, and continuous guidance throughout my life, this work would never have been done. Next comes my parents, the best gift I have ever received in life. The great father and my inspirer Doctor Nahar, and the soulful mother, my angle, Abla. Without their satisfaction and prayers, I would not be who I am today. My gratitude is extended to my lovely wife Rodab for her support and great patience at all times, sure not to forget my beloved daughter Lamar in spite of the mess she used to do at my home-office everyday!

In this opportunity, I would like to thank all those who assisted me during my Ph.D. studies. First and foremost, I would like to thank my supervisor Professor Zbigniew Dziong for the support, valuable advices, and help. Professor Dziong always provided an open door to his office, for which I am really grateful. Working with him added a valuable experience to my life, which really deserves my full acknowledgements.

I would like to thank Dr. Hadi Otrok for the advices and valuable recommendations. Special thanks for my examiner committee: our guest Professor Anjali Agarwal, Professor Michel Kadoch, and the president Professor Witold Suryn. Thanks for giving me the honor by being my Ph.D. committee, and for your valuable time and efforts.

Thanks to my colleagues Dr. AbdAllah Jarray, Dr. Con Tran, Rzia Mossanen, and all members of the LAGRIT Lab.

Last, but by no means least, I would like to acknowledge the financial support, through my Supervisor Professor Dziong, to Bell Canada and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

RESOURCE MANAGEMENT FOR VIRTUALIZED NETWORKS

Ahmad Quttoum

RÉSUMÉ

La virtualisation des ressources a émergé comme une approche prometteuse qui peut être employée pour améliorer l'efficacité des technologies de gestion des ressources dans un réseau. Ainsi, dans ce travail, l'objectif est d'étudier comment automatiser la gestion de la bande passante dans un réseau, tout en utilisant un schéma de partitionnement virtuel des ressources. Dans la littérature plusieurs travaux de recherche ont abordé la gestion des ressources dans le contexte des réseaux virtuels privés, mais chacun a ses limites. Un partage non équitable des ressources, une faible utilisation de la bande passante disponible, faibles profits, l'exagération, et la collusion sont tous des exemples de ces limitations. En effet, l'absence des systèmes capables de garantir un partage équitable de la bande passante entre les utilisateurs, encourage l'exagération en besoin des ressources. C'est-à-dire, un client peut monopoliser les ressources qui sont censées servir les autres utilisateurs. Un partitionnement statique des ressources peut limiter le problème d'exagération, toutefois il peut en résulter une mauvaise utilisation de la bande passante, ce qui signifie des profits moindres pour les fournisseurs de services réseau virtuel. Cependant, le déploiement d'une technologie de gestion de bande passante autonome peut améliorer l'utilisation des ressources, et maximiser le taux de satisfaction des clients. Une telle technologie peut fournit également aux clients une sorte de privilège qui peut être utiliser pour tricher. Ainsi, des actions de tricherie comme la collusion et l'exagération peuvent être adopter par certains clients. Résoudre des telles problèmes est aussi abordé dans ce travail.

Dans la première partie de cette thèse, nous nous proposons de résoudre les problèmes de la mauvaise utilisation de la bande passante, des faibles profits pour les fournisseurs de services et les taux de blocage élevés rencontré dans l'approche traditionnelle "Premier arrivé premier servie " (FAFA) . La solution proposée est basée sur un mécanisme de gestion des ressources Autonome (ARMM). Cette solution déploie un algorithme intelligent d'allocation de la bande passante basé sur le mécanisme de vente aux enchères. Pour réduire la tendance de l'exagération, nous proposons d'utiliser la technique Vickrey-Clarke-Groves (VCG), qui consiste en un modèle de menace qui pénalise les clients qui exagèrent leurs besoins en bande passante. La pénalité pour chaque client qui exagère son besoin en bande passante est calculée en se basant sur les désagréments qu'il cause à autrui dans le système. Pour résister à la collusion entre clients, nous proposons de calculer un prix fictif seuil dépendant de l'état du système et en se basant sur la théorie de décision de Markov. En effet, nous calculons pour chaque état du système un seuil de prix de vente pour chaque unité de la bande passante disponible dans le réseau. Tout client offrant un prix de vente par unité de bande passante inférieure au seuil calculé est automatiquement bloqué et ne peut utiliser les ressources du réseau.

Dans la deuxième partie de cette thèse, nous nous proposons de résoudre une version élargie du problème d'allocation de la bande passant dans un réseau virtuel privé, et en utilisant une

méthodologie différente. Le problème d'allocation de la bande passante est étendu à un problème de partitionnement. Une telle expansion permet de calculer la répartition optimale de la bande passante de chacun des liens du réseau entre les différentes classes de service et de fournir une meilleure utilisation des ressources du réseau. Afin de trouver les paramètres d'une gestion optimale des ressources d'un réseau virtuel privé, le problème est résolu par les techniques de la programmation linéaire (LP). Un schéma de partitionnement dynamique de bande passante est également proposé pour surmonter les problèmes liés au partitionnement statique des ressources telle que un faible taux d'utilisation de la bande passante disponible, qui peut résulter en des partitions sous-utilisées. Ce modèle de partitionnement dynamique est déployé d'une manière périodique. Un partitionnement périodique fournit une nouvelle façon de réduire le raisonnement de l'exagération, par rapport au modèle basé sur la menace de payer une pénalité relative au désagrément posé à autrui. Nous réduisons ainsi le coût lié au calcul de cette pénalité.

Dans la troisième partie de cette thèse, nous proposons un système de gestion décentralisé pour résoudre les problèmes mentionnés ci-dessus dans le contexte des réseaux virtuels privés gérés par des opérateurs des réseaux (VNOs).Cette décentralisation permet le déploiement d'un niveau supérieur de la gestion autonome des ressources du réseau, à travers lequel, les responsabilités de gestion sont répartis sur les nuds du réseau, chacun étant responsable de la gestion de ses liens sortants. Comparé aux architectures centralisées, une telle distribution permet une plus grande fiabilité et un dimensionnement de la bande passante plus facile et plus efficace. En outre, il crée un cadre de la concurrence à deux faces qui permet une double vente aux enchères parmi les acteurs du réseau, les clients et les contrôleurs des nuds du réseau. Un tel environnement de compétition offre une nouvelle façon de réduire l'exagération comparée au modèle périodique basé sur la menace de payer une pénalité. Mieux encore, ce nouveau modèle peut offrir des meilleurs taux d'utilisation, un taux de blocage moins élevé et par conséquent des profits pour le fournisseur de service plus importants. Enfin, des simulations numériques et des résultats empiriques sont présentés pour soutenir les solutions proposées, et fournir une comparaison plus éclairée avec d'autres modèles qui ont été proposés dans la littérature.

Mot-clés: Réseaux Virtuels, de gestion autonome, partitionnement virtuel, gestion des ressources de bande passante, la théorie des jeux, Mechanism Design, Théorie de la décision de Markov, et de la Programmation Linéaire.

RESOURCE MANAGEMENT FOR VIRTUALIZED NETWORKS

Ahmad Quttoum

ABSTRACT

Network Virtualization has emerged as a promising approach that can be employed to efficiently enhance the resource management technologies. In this work, the goal is to study how to automate the *bandwidth* resource management, while deploying a virtual partitioning scheme for the network bandwidth resources. Works that addressed the resource management in Virtual Networks are many, however, each has some limitations. Resource overwhelming, poor bandwidth utilization, low profits, exaggeration, and collusion are types of such sort of limitations. Indeed, the lack of adequate bandwidth allocation schemes encourages resource overwhelming, where one customer may overwhelm the resources that supposed to serve others. Static resource partitioning can resist overwhelming but at the same time it may result in poor bandwidth utilization, which means less profit rates for the Internet Service Providers (ISPs). However, deploying the technology of *autonomic management* can enhance the resource utilization, and maximize the customers' satisfaction rates. It also provides the customers with a kind of privilege that should be somehow controlled as customers, always eager to maximize their payoffs, can use such a privilege to cheat. Hence, cheating actions like exaggeration and collusion can be expected. Solving the aforementioned limitations is addressed in this work.

In the first part, the work deals with overcoming the problems of low profits, poor utilization, and high blocking ratios of the traditional First Ask First Allocate (FAFA) algorithm. The proposed solution is based on an Autonomic Resource Management Mechanism (ARMM). This solution deploys a smarter allocation algorithm based on the *auction mechanism*. At this level, to reduce the tendency of exaggeration, the Vickrey-Clarke-Groves (VCG) is proposed to provide a threat model that penalizes the exaggerating customers, based on the inconvenience they cause to others in the system. To resist the collusion, the state-dependent shadow price is calculated, based on the Markov decision theory, to represent a selling price threshold for the bandwidth units at a given state.

Part two of the work solves an expanded version of the bandwidth allocation problem, but through a different methodology. In this part, the bandwidth allocation problem is expanded to a bandwidth *partitioning* problem. Such expansion allows dividing the link's bandwidth resources based on the provided Quality of Service (QoS) classes, which provides better bandwidth utilization. In order to find the *optimal* management metrics, the problem is solved through Linear Programming (LP). A *dynamic* bandwidth partitioning scheme is also proposed to overcome the problems related to the static partitioning schemes, such as the poor bandwidth utilization, which can result in having under-utilized partitions. This dynamic partitioning model is deployed in a *periodic* manner. Periodic partitioning provides a new way to

reduce the reasoning of exaggeration, when compared to the threat model, and eliminates the need of the further computational overhead.

The third part of this work proposes a *decentralized* management scheme to solve aforementioned problems in the context of networks that are managed by Virtual Network Operators (VNOs). Such decentralization allows deploying a *higher level* of autonomic management, through which, the management responsibilities are distributed over the network nodes, each responsible for managing its outgoing links. Compared to the *centralized* schemes, such distribution provides higher reliability and easier bandwidth dimensioning. Moreover, it creates a form of *two-sided competition framework* that allows a *double-auction* environment among the network players, both customers and node controllers. Such competing environment provides a *new way* to reduce the exaggeration beside the *periodic* and *threat* models mentioned before. More important, it can deliver better utilization rates, lower blocking, and consequently higher profits.

Finally, numerical experiments and empirical results are presented to support the proposed solutions, and to provide a comparison with other works from the literature.

Keywords: Virtualized Networks, Autonomic Management, Virtual Partitioning, Bandwidth Resource Management, Game Theory, Mechanism Design, Markov Decision Theory, and Linear Programming.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ARB	Autonomic Resource Broker.
ARMM	Autonomic Resource Management Mechanism.
ASA	Autonomic Service Architecture.
BFS	Breadth First Search.
BR	Bandwidth Borrowing.
CAC	Connection Admission Control.
CDRP	Centralized Dynamic Resource Partitioning.
CP	Complete Partitioning.
CS	Complete Sharing.
DDRP	Decentralized Dynamic Resource Partitioning.
DP	Dynamic Partitioning.
FAFA	First Ask First Allocate.
GoS	Grade of Service.
HP	Hewlett-Packard.
ILP	Integer Linear Programming.
IP	Internet Protocol.
IPSec	Internet Protocol Security.
IPTV	Internet Protocol Television.
ISP	Internet Service Provider.

IT	Information Technology.
ITU	International Telecommunication Union.
LAN	Local Area Network.
LP	Linear Programming.
LSP	Label Switched Path.
MA	Management Agent.
MDP	Markov Decision Process.
MDPD	Markov Decision Process Decomposed.
MIDAS	Multi-unit Dutch Auctions.
MPLS	Multi-Protocol Label Switching.
NA	Node Agent.
NE	Nash Equilibrium.
NP	Nondeterministic Problem.
PF	Price Freezing.
PPP	Profit Percentage Parameter.
PSP	Progressive Second Price.
QoS	Quality of Service.
RSVP-TE	Resource Reservation Protocol - Traffic Engineering.
SCF	Social Choice Function.
SLA	Service Level Agreement.

SON	Service Overlay Network.
VCG	Vickrey Clarke Groves.
VN	Virtualized Network.
VNO	Virtual Network Operator.
VoIP	Voice over Internet Protocol.
VP	Virtual Partitioning.
VPN	Virtual Private Network.
VS	Virtual Service.
WiFi	Wireless Fidelity.
WiMAX	Worldwide Inter-operability for Microwave Access.

INTRODUCTION

Nowadays, the world is living through a huge Internet application explosion era that imposes a great management challenge to the Internet Service Providers (ISPs). Indeed, ISPs are expected to satisfy different levels of Quality of Service (QoS) requirements that may vary from one customer to another. Handling such challenge requires new management techniques that can ease the process of management at the ISPs' side. Currently, most service management models require direct intervention from the ISPs, Kim *et al.* (2005). This leads to slow response time and high operational costs. More important, such models can deliver high rates of customers' dissatisfaction, and increase the management cost at the providers' side, which means less profits.

Network management started to become a challenging and time-consuming business for human network managers, however, the notion of *Network Virtualization* has emerged as a promising method for efficient network management, Niebert *et al.* (2008). In the literature, it is defined as the method of splitting the network physical resources into separate *virtual* (or logical) portions, each of which can be assigned to a particular customer (or a group of customers), Rixner (2008), Chowdhury and Boutaba (2010). These virtual portions can be managed by the network customers in an *autonomic* way. Hence, a *Virtualized Network* (VN) is a network where the available resources (e.g. the bandwidth) is logically divided among its customers. Each customer is allocated a sufficient amount of resources that accommodate its needs and satisfy its QoS requirements. In such VN, customers are responsible for self-manage and control their allocated resources.

The virtual partitioning concept is not new. However, combining the *virtual partitioning* with the *autonomic management* brought a new aspect to the theme of VNs, which promise to provide an efficient methodology for the future management systems. Such an approach is intended to provide an improved efficiency, satisfaction, and performance to the network administrators. Moreover, it reduces the management complexity by distributing many tasks among its customers.

Figure 0.1 illustrates the scheme of bandwidth provisioning in VNs, and shows how the concepts of *virtual partitioning* and *autonomic management* are applied. In the figure, the bottom part shows an example of an IP network that consists of several nodes connected together through a group of links with specific amount of bandwidth resources. The middle part represents an abstracted level of the underlying network that is called a root VN.

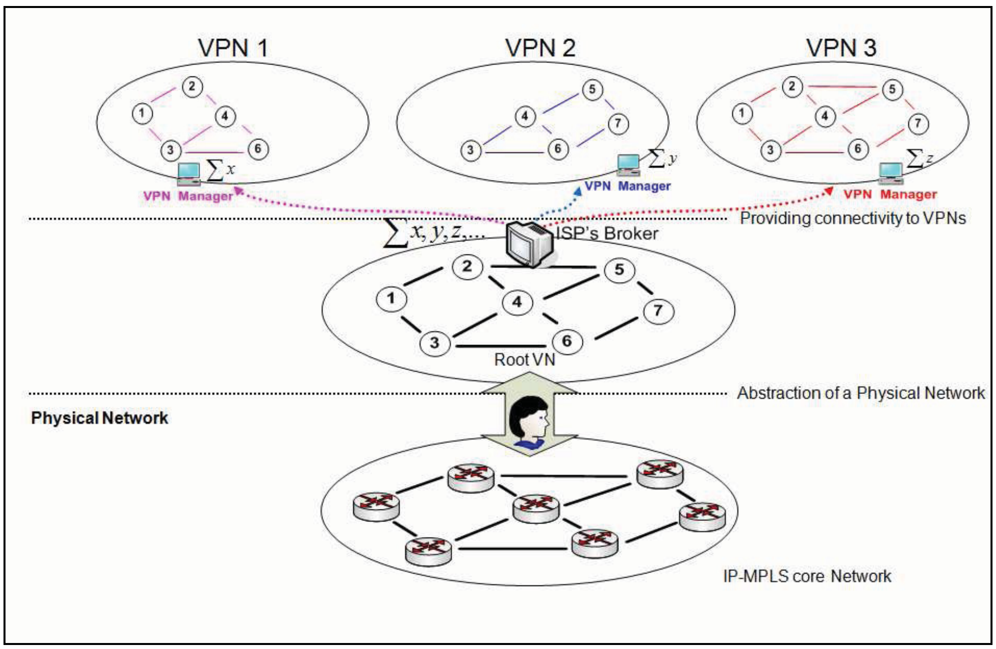


Figure 0.1 Provisioning in Virtualized Networks

The top part of the figure shows an example of how the virtual partitioning and autonomic management are deployed. In the context of virtual partitioning, the bandwidth resources of the root VN are partitioned among the three new Virtual Private Network (VPNs). A VPN could be defined as a group of computer systems that are connected together through a public network infrastructure as a private network, by deploying the concept of virtual partitioning, Cohen and Kaempfer (2000). Hence, each VPN has a portion of resources over some network links, through which, it can serve its connection demands. In autonomic management, for each VPN, the roots VN assigns a VPN manager that is responsible of self-manage and control the underlying network using its virtual portion of resources.

0.1 Motivation

Since the last couple of decades, telecommunication needs among the business fields and others passed through a huge era of globalization. Instead of simply dealing with local or regional concerns, many started to think about global markets. Several companies have facilities that spread around the world, and hence they all started to look for a way that maintains fast, secure, and reliable communications wherever their offices are. Until fairly recently, this meant leasing *dedicated private lines* to maintain private networks.

Dedicated leased lines provided companies with a way to expand their private networks beyond its local geographic area. Such private networks have obvious advantages over the public networks when it comes to reliability, performance, and security. However, maintaining such networks, particularly when using dedicated physical lines, can be *quite expensive* and often the cost becomes prohibitive for large distance between the offices.

As the popularity of the Internet grew, business people turned to using Internet in order to extend their own private networks. First came Intranets, which are password-protected sites designed to be used only by the company employees, Stenmark (2002). Nowadays, many companies are leasing their own VPNs to accommodate the needs of remote employees and distant offices.

As depicted in Figure 0.2, VPNs can be deployed to connect remote and geographically distributed sites and location to act as a single network. Usually, such VPNs are built on top of a public network infrastructure (i.e. the Internet). Being a new technique to replace the dedicated private networks, these VPNs must provide comparable services in terms of QoS, security, reliability, and cost-effectiveness.

In recent years, a substantial progress in the technologies of IP security has enabled the existing VPNs to provide customers with sufficient levels of secure and private communication systems, compared to that offered by the dedicated line technologies (Frame Relays, IPSec), Duffield *et al.* (2002). The same progress had been achieved in the fields of routing and tunneling issues, while much less attention has been made to the *resource management* fields in VPNs.

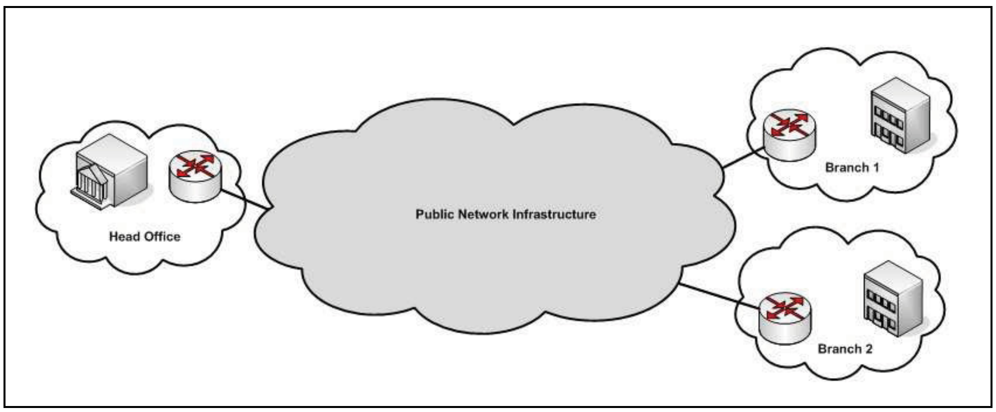


Figure 0.2 Virtual Private Network

According to such increased progress in security and the overwhelming success of IP networking technologies, the number of VPNs' users is growing rapidly, and it emerged as a leading solution for multi-site enterprise communication demands. Therefore, there is a need to find an efficient management scheme that can support such variety of VPNs' users.

Deploying the concept of *autonomic management* for such VPNs brought a new dimension to the desired management models that can attain greater degrees of scalability, and manage more differentiated applications than in traditional networks.

VNs concept still requires solving many research issues and challenges. Hence, combining the concepts of autonomic management and virtual partitioning need to be studied in order to determine the right trade-off between such virtualized isolation and the desired performance.

0.2 Problem Overview

To demonstrate the addressed problem better, let us consider the sharing of the network resources among the three VPNs illustrated in Figure 0.1. In this context, the term resources refer to the *bandwidth* allocated to the links, which is used to carry connections between the network nodes. Obviously, there are many links shared between the three VPNs (i.e. the links connecting the node couples 3-4, 4-6, and 3-6) and in such allocation scenarios the following problems can be encountered:

0.2.1 Poor Bandwidth Utilization

Deploying the *virtual partitioning* concept means that these three VPNs have virtual portions of the bandwidth resources over the considered links. Each VPN might have different traffic pattern that changes in a time-dependent manner due to the varying types of the provided services, types of customers, and the different business hours. Consequently, such virtual partitioning of the network bandwidth resources should consider the dynamic (real-time) customers' requirement. From the ISPs' perspective, relying on a *static* resource partitioning scheme is not satisfying, since this may waste the network resources, and reduce their total profits. Indeed, since static partitioning decisions do not take into account the varying traffic demands of the network customers, some portions of resources might be overloaded while others underloaded, resulting in poor utilization rates of the network resources and non-satisfying QoS guarantees.

0.2.2 Cheating Actions

Deploying the *autonomic management* concept gives the VPN operators a kind of privilege that enables them to self-manage and control their network resources. Hence, the provisioning process is done according to the revealed requirements of the VPN operators. From the ISPs' perspective, giving the VPN operators such a privilege can provide better utilization of the network resources, assuming that the VPN operators know their changing requirements better, and accordingly they can acquire the required resources based on their actual needs. Naturally, as long as the VPN operators behave according to their real needs, deploying such an approach can increase bandwidth utilization due to the statistical multiplexing, which results in better satisfaction rates and higher profits at the same time. However, VPN operators may try to maximize their own payoffs, by executing some cheating actions facilitated by the autonomic management. As an example, VPN operators may tend to *exaggerate* their requirements and ask for extra resources, such behavior can waste the network resources and increase the blocking ratios for those non-exaggerating operators. Exaggeration can be motivated by many reasons like the ones listed below:

- usually, more bandwidth means higher transmission quality. Therefore, VPN operators might tend to obtain the largest possible amount of bandwidth resources for their transmissions, even if the resulting quality improvement is minimal;
- to cope with any sudden changes in the network state, VPN operators may tend to always keep an extra amount of resources. This point is also motivated by the fact that the static bandwidth partitioning does not provide QoS guarantees for unexpected traffic surges;
- the lack of resource utilization incentives can also motivate exaggeration. Having no incentives, VPN operators are not expected to smartly route their connections. However, they may rely on their ability to resend any delayed or lost connections using the extra resources they have.

Moreover, deploying an auction mechanism for resource allocation may open the way for another cheating action like *collusion*. In such case, auctioneers may collude together and bid a single low price in order to force the seller (resource provider) to sell the resources at low prices. Collusive bids can serve the auctioneers by maximizing their payoffs. On the other hand, such actions may sharply minimize the providers' profits.

0.3 Objectives

The main goal behind this work is to solve the problems discussed in the previous section. Hence, the first main objective is to study and develop efficient resource utilization models that enable the ISPs to attempt a dynamic bandwidth allocation mechanism instead of the static one.

This dynamic model should allocate the network bandwidth resources based on the customers' real-time requirements. Besides, such model should also satisfy the goals of maximizing the ISPs' profit, while maintaining the QoS guarantees of the VPN operators.

The second main objective is to study and develop models that motivate the VPN operators to behave cooperatively, in a way that serves the whole network interest and therefore also avoids cheating actions like exaggeration and collusion.

The resulting work should answer the following questions:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to suppress collusion actions that harm the allocation process?
- How to measure the optimal bandwidth division among QoS classes?
- How to decentralize the resource allocation mechanism?

These questions will be answered by developing different models using different theories and working environments. However, achievements from the proposed models will be compared to others provided in the literature.

0.4 Thesis Outline

The thesis is organized as follows: Chapter 1 is divided into two parts, Section 1.1 provides an overview of the tools and theories employed to solve the aforementioned problems and satisfy the work objectives. This overview targets the Game theory, Mechanism Design, Markov decision theory, and the Linear programming theory. Section 1.2 presents a literature review related to the addressed problem, and explains how the chosen tools and theories were deployed in the previous works. Then Chapters 2, 3, and 4 show the contribution of this work.

Chapter 2 studies the problem of resource management, in which it mainly discusses the issue of *poor bandwidth allocation* and the problems of users' cheating actions. Here, the work addresses a problem that is limited to management of resources belonging to one link. Such network link has some bandwidth resources to be allocated for different VPN operators. The proposed solution deploys the concepts of *auction mechanism*, and *autonomic management*. Deploying the auction mechanism presented in Section 2.4.3 provided a smart way to set up an efficient bandwidth allocation model. Through the autonomic management, the system authorizes the VPN operators to self-manage and control their allocated resources. However, in such cases, cheating actions are profitable for users. In particular, having the privilege of autonomic management, *exaggeration* is expected since users may ask for more resource than their real needs, which wastes the network resources and minimizes the providers' profits. To reduce the motivations of exaggeration, in Section 2.5, the Vickrey-Clarke-Groves (VCG) mechanism from the *Game Theory* is applied to enforce cooperation in such non-cooperative scenario. Using the VCG, the model penalizes the exaggerating users based on a threat model that measures the inconvenience each user cause to the whole users in the network. Moreover, under such auction scenarios, *collusion* is another possible way of cheating the users may follow to maximize their profits. To resist the collusion, in Section 2.6, from the *Markov Decision Theory*, a minimum selling price value is calculated based on the state-dependent shadow price concept. The calculation of this value is derived in Section 2.6.2. Hence, it represents a dynamic selling price threshold that is used to define whether to accept or reject the users' offered bids. Compared to the First Ask First Allocate (FAFA) bandwidth allocation algorithm, results in Sections 2.7 and 2.8 show that this work provides better use of the bandwidth resources, resulting in higher profits and lower blocking ratios. Moreover, it delivers an efficient method to reduce the effect of the *exaggeration* and *collusion* cheating actions.

Recalling the list of objectives presented in Section 0.3, the contribution of this part lies in answering the following question:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to suppress collusion actions that harm the allocation process?

In terms of novelty, deploying the threat model presented in Section 2.5 provides a novel way to suppress exaggeration, and enforce fair resource allocations in VPNs. Moreover, although the VCG truth-telling Mechanism Design was applied in the fields of wireless networks, in this work, it is applied in the wired VPNs. The Markov Decision Process (MDP) was recently deployed in the fields of Service Overlay Networks (SON) for capacity adaptation, however, in this work it is deployed to find the optimal selling price threshold for the ISP, which is a novel work.

Chapter 3 solves the same resource management problem, but for a different system. Compared to the previous part of the work, the system is expanded from a single link to a full network problem. Hence, this part adds the issue of path selection to the problem. The presented work mainly deals with the problem of how to *utilize the bandwidth resources better*, and how to provide *optimal* management metrics. To address the issue of different QoS requirements in a better way, this part also considers *bandwidth partitioning* between different QoS classes. In addition, Section 3.4.1.1 shows how this bandwidth partitioning scheme is deployed in a *dynamic* manner, which provides a more efficient solution to the static scheme presented in Section 3.4.2. Moreover, to reduce the exaggeration problem, the resource partitioning process is executed dynamically in a *periodic* manner. Such periodic scenario provides a new way to reduce the tendency of exaggeration, without the need of the threat model proposed in Chapter 2. As presented in Section 3.4.1.2, the problem here is solved using the *Linear Programming Theory* that provides optimal metrics. Compared to other models in the literature, Section 3.6 shows how such optimal metrics result in better resource utilization, lower blocking ratios, and higher profit rates.

From the list of Section 0.3, this part answers the following question:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to measure the optimal bandwidth division among QoS classes?

The novelty of this part lies in deploying the Linear Programming to solve autonomic resource management problems, with dynamic bandwidth allocation schemes for VPNs.

In Chapter 4, the work addresses the problem of resource management in networks that are managed by Virtual Network Operators (VNOs). In this case the network leases its bandwidth resources from different ISPs. Such network is managed by a VNO that represents a virtual ISP, and it aims to utilize these bandwidth resources optimally in order to gain higher profits. Compared to the previous two chapters, the solution in this chapter takes a different management direction, where: (1) The model is developed to solve the resource management problem in a *decentralized* way, in which, as shown in Section 4.4, the VNO distributes the management load over special entities that manage the network nodes. These entities, named Node Agents(NAs), are proposed to *reduce the management load* at the controller side according to Algorithm 4.1. (2) The model deploys *a higher level of autonomic management*, where not only the network users, but also the NAs are responsible to self-manage and control their resources. So, the model here deploys a two-sided autonomic management scheme. At each side, if cooperation with the network controller is guaranteed, such decentralization scheme can provide *higher reliability* and *easier network dimensioning*. (3) To enforce cooperation, as depicted in Figure 4.2, a *double-auction* framework is proposed to create a competition environment between the NAs (resource sellers) from one side, and the users (buyers) from the other side. In this way, NAs will be motivated to use their available resources better (*this means better utilization in the network*) in order to offer competing selling prices to win the users' bids (*which means lower blocking ratios in the network*). On the contrary, network users will be competing to offer higher bids (*higher total profits from the network*) to win the resource allocations from the NAs. (4) Although this model is deployed in a periodic manner that reduces exaggeration, still, the *two-sided competing environment* represents another way to reduce the exaggeration. With such environment, exaggerating users will not be able to offer competing high bids, same way, exaggerating (or monopolizing) NAs will not be able to offer low selling prices to attract the buying users.

This model is solved through Linear Programming. In a hierarchical manner, Section 4.4.3 shows the NAs' model, which is solved by each NA in a decentralized way. In contrast, Section 4.4.2.3 shows the central model that is solved by network controller. Thus, results of the decentralized functions provide the input of the central one.

Compared to other central management scenarios, this part of the work provided different methodologies from that of the former parts to answer the first three objective questions. Indeed, the decentralized management model along with the double-auction scenario provide better resource utilization, which implicitly means *higher profit and satisfaction ratios*. Also, the two-sided competing environment represents *another new way* to reduce the motivations of exaggeration.

Thus, the achievements of this part answer the following:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to measure the optimal bandwidth division among QoS classes?
- How to decentralize the resource allocation mechanism?

In terms of novelty, deploying the Linear Programming theory to solve a decentralized dynamic resource management problems with double-auction framework in VPNs is a novel work, where previous works addressed static centralized allocation schemes only.

Finally, the thesis ends by a conclusion that provides a summary of the addressed problem and its proposed solutions. It also shows the achieved objectives, the work novelty, a proposition for future works, and deliverable publications.

0.5 Summary of Achievements and Novelty

The work of Chapter 2 provided better usage of the bandwidth resources, compared to the FAFA bandwidth allocation algorithm. This resulted in *higher profits* and *lower blocking ratios*. Moreover, it delivered efficient methods to reduce the effect of the *exaggeration* and *collusion* cheating actions.

In Chapter 3, the work provided new techniques for optimal bandwidth partitioning. Such techniques helped in *utilizing the network bandwidth resource better*, gain *higher profits* with *lower blocking ratios*. Also, this part of the work provided different methodologies from that of the former part to solve the problems of exaggeration and bandwidth utilization. Indeed, it reduced the *exaggeration* motivations through the *periodic auctions* instead of the threat model used in Chapter 2. For resource utilization, this part employed the Linear Programming to solve an *optimal* dynamic resource partitioning model.

The decentralized model of Chapter 4 proposed a new model based on a *higher level* of autonomic management, with a *double-auction* scenario. Through which, it provided *higher bandwidth utilization rates* that implicitly means *higher profit and satisfaction ratios*. The double-auction allowed internal competition within the NAs and the network customers. Among NAs, this resulted in *better utilization* and *lower blocking ratios*, while it delivered *higher total profits* from the competition among the network customers. Moreover, in addition to the reduction of the *exaggeration* motivations due to the periodic allocations, in this part, the two-sided competing environment represented *another new way* to reduce exaggeration.

In summary the main novelty of the proposed work consists of the following elements:

First

Suppressing exaggeration and enforcing fairness resource allocations in VPNs based on a threat model.

Second

VCG truth-telling Mechanism Design has never been applied in wired VPNs. However, it is applied in the fields of wireless networks.

Third

Suppressing cheating and collusive behaviors in VPNs.

Forth

Markov Decision Process (MDP) was recently deployed in the fields of Service Overlay Networks (SON) for capacity adaptation. In this work, it is deployed to find the optimal selling-threshold price for ISP.

Fifth

The use of Linear Programming in autonomic resource management, with dynamic bandwidth allocation schemes for VPNs.

Sixth

Deploying dynamic decentralized resource management with Linear Programming in VPNs is a novel work. Previous works attempted static centralized schemes.

CHAPTER 1

METHODOLOGY TOOLS AND LITERATURE REVIEW

This Chapter provides a presentation to the tools and theories used to achieve the aforementioned objectives, and the related work in the literature. Hence, and to make this chapter self contained, Section 1.1 discusses the theories and techniques employed to achieve the desired objectives. In turn, Section 1.2 goes through the work done in the literature, such work that generally deals with the addressed problem.

1.1 Methodology Tools

This section lists brief high-level definitions of the proposed tools and theories to be used in this work.

1.1.1 Game Theory

The Game theory could be defined as a mathematical tool that analyzes the strategic interactions among a group of rational decision makers, Myerson (1991). Through the last few decades, Game theory shows the ability to provide satisfactory results in varying disciplines. Those that include politics, management, economics, social science, psychology, and computer science. To accomplish a game, there are three major components that must be clearly defined in a strategic-form, Morris (1994), these components are: (1) The set of players. (2) The strategy (i.e. action space) of each player. (3) The payoff function that measures the outcome of the game for each player.

In Wang *et al.* (2008) and others, the authors present the Game theory as a powerful tool that can be employed to study the interaction scenarios among multi-player decision problems. These types of problems where players may have different objectives to satisfy, and hence, having different strategies of behavior. Thus, competition and other interaction scenarios are expected. In this context, results may represent positive gains for some players, and negative ones for others.

By breaking-down its definition, the following main characteristics are noticed:

- strategic: this implies the behavior of the game player, where each player has a set of candidate behaviors to choose from. The chosen strategy affects the expected outcomes of the game;
- interaction: represents the effect that a player's strategy impose to the other players in the game. Such interaction can affect one other player at least;
- group: it refers to the decision makers in the game, usually they are known as the game players and referred to as a group;
- rational: the player who aims to maximize its own payoff is called rational player. Hence, a rational player usually chooses its strategy based on the other players' chosen strategies.

In Game theory, the strategic interaction among the players is modeled in a formalized incentive structure. So, it not only provides models for efficient self-enforcing distributed design, but also, derives well defined equilibrium criteria to study the optimality of the game outcomes in various scenarios, i.e., static or repeated, cooperative or non-cooperative, Morris (1994). Cooperative games are those where multi-players all cooperate together in order to achieve the maximum possible payoffs for all, Barron (2007). On the other hand, non-cooperative games are those where the game players compete with each other, and every player cares only about its own payoff, Morris (1994). This is different from the selfish behavior, where the players care only about maximizing their own goal, even if this causes problems to others, Wang *et al.* (2008).

Usually, the outcome of a non-cooperative game is given by the Nash Equilibrium (NE) point, Nash (1951). This NE could be defined as the set of strategies for all players, such that no player can improve its own utility by unilaterally deviating from the equilibrium strategy, given that all other players adopt the equilibrium strategies. So, the NE indicates that no individual player would have the incentive to choose a different strategy.

Static non-cooperative games are played only once, where the players are myopic and only care about their current utility, Morris (1994). The competition between selfish players in static games often results in an NE that is not system efficient. Therefore, stimulation of cooperation among selfish players is very important in order to achieve social welfare. However, repeated non-cooperative game models may capture better interactions in long-run scenarios, Wang *et al.* (2007).

In repeated game modeling, the players play the similar static game many times, so their future decisions are conditioned on other players' past moves. In this way, one may expect that cooperation can be enforced by establishing the threats of punishment, individual reputation, mutual trust, and so on.

1.1.2 Mechanism Design

Mechanism Design could be considered as a sub-field of the Game theory and microeconomics, Mas-Colell *et al.* (1995). By employing the Game theoretic tools and its strategic analysis, with the Mechanism Design, one can achieve the desired goals. Hence, the Game theory can be used to study and analyze the outcomes of the players' expected behavior. On the contrary, the Mechanism Design helps in defining rules based on a Social Choice Function (SCF) to restrict the players' behavior in a way that serves the whole interest, Wang *et al.* (2008). Following such rules, players can gain positive payoffs, otherwise they will be subjected to certain consequences. The Mechanism Design has been successfully deployed in various disciplines like resource allocation, auctions, electronic markets, computer science, and many others.

Being a mechanism that studies the solutions of private information games, the Mechanism Design is employed to provide optimal system-wide solutions for distributed optimization problems, with self-interested players. In this context, the mechanism designers are allowed to choose the game rules that, according to their SCF, govern social welfare for all. Thus, the designers are usually interested by the game's outcome represented by the NE point. Rules should be designed in a way that leads the players to follow the dominant strategy that delivers the SCF.

1.1.2.1 Vickrey-Clarke-Groves Mechanism

Briefly, the Vickrey-Clarke-Groves (VCG) is a strategy-proof mechanism that has been proposed by Vickrey (1961), Clarke (1971), and Groves (1973). VCG is considered as the only allocation-efficient, and strategy proof mechanism amongst all direct revelation mechanisms. As an example, VCG is used to solve the allocation problems based on auctions, where it follows the sealed-bid type of auctions. These auctions, in which, the players submit different bids for different items. Having a social welfare decision, the system allocates the resources in a way that maximizes the total utility of the system. In VCG, the mechanism ensures that for each bidder, the dominant strategy is to truthfully reveal the real valuation of the item. This results in maximizing the total valuations of the auction items, which represents gain maximization for the sellers.

1.1.3 Markov Decision Theory

The Markov decision theory could be defined as a tool that provides efficient decision making frameworks based on mathematical modeling, Puterman (1994). Such theory has been successfully deployed in varying disciplines, including economics, computer science, inventory control, robotics, etc. The Markov decision theory is considered as an outgrowth of the Markov process theory and the Dynamic programming, Dziong (1997). The Dynamic programming concept was first introduced by Bellman (1957). Combining the Dynamic programming with the Markov process theory was proposed by Howard (1960). This combination results in the development of the policy-iteration algorithm that is used to calculate the optimal decision policies. Hence, the Markov decision theory has emerged as a powerful tool that can analyze probabilistic sequential decision processes with an infinite planning horizon, Dziong (1997).

According to Dziong (1997), Markov processes, or Markov chains as usually called, are based on two fundamental concepts: states and state transitions. In this context, the state can be considered as a random value that describes the system at a given time instant, e.g., the number of connections over a network link. The state transition describes the dynamic changes in the system state at a given time instant, in the given example, changes may occur by new connection arrival, connection delivery, or connection drop. Generally, Markov processes can be classified in either discrete or continuous time-spaces. This depends on the shape of the system review instances, wether it happens at fixed or random times, respectively. These state transition instances can be controlled by the system controller, i.e., the ISP in the given example. In such scenarios, the ISP can define what would be the optimal control decision, i.e., the optimal shape of the state transition.

A prescription for the final decision in each time instance is called policy. Derived from the Markov decision theory, the notation of *shadow price* can be used to calculate the optimal policy. In the given example, the state-dependent shadow price can represent the dynamic cost of admitting a connection request over certain bandwidth unit, at a given state. This shadow-price notation was developed for Call Admission Control (CAC) and routing in telecommunication networks. As shown in Dziong (1997), the state-dependent shadow price can be calculated using efficient methods like the policy-iteration algorithm, and the value-iteration algorithm.

It is worth to mention that the notion of average link shadow price was first introduced by Kelly (1988) for optimization of adaptive load sharing. In Dziong *et al.* (1988), the authors introduced the state-dependent link shadow price notion that was used for optimization of state-dependent routing.

1.1.4 Linear Programming Theory

Linear Programming could be defined as a mathematical method that can be employed by the means of achieving optimal outcomes. Usually, it finds the minimum or the maximum possible value for a problem modeled through mathematical linear relationships bounded by certain constraints, Dantzig (1963), and Williams (1965). However, in this context, it is worth to mention that the word *programming* refers more to the word *planning* rather than its popular indication. Linear Programming has successfully proved to deliver optimal metric in various field that include economics, energy, manufacturing, routing, etc. In fact, Linear Programming has emerged as powerful tool that can be deployed almost in everything, from airline scheduling to simple pricing problems in trades and distribution.

According to Chinneck (2001), for *constrained optimization* problems, Linear Programming could be considered as the most-widely applied optimization theory. Compared to the unconstrained problems, solving a constrained optimization problem is more challenging, where it adds more restrictions to the desired optimality point. Hence, finding the optimal solution in a constrained problem does not necessarily mean the points of the peak or the valley, it could be at somewhere in between bounded by the objective functions' constraints.

Building up a constrained optimization model, the following four element should be clearly defined, Chinneck (2001).

- objective function: This element represents the problem to be optimized, either maximized or minimized. For example, the objective might be profit maximization, or capital-cost minimization. Mathematically, it comes in the form of equation that combines the model variables;
- constraints: These elements represent the limits of the possible outcome from the objective function, similarly they combine the variables that express the anticipated objective of the model;
- decision variables: These variables are those optimized by the model, i.e., the decisions that maximize or minimize the value of the objective function. Hence, their values are not known at the beginning, but while solving the problem, they are continuously adapted till the optimal value of the objective function is obtained;
- variable bounds: These represent the minimum and maximum values of the decision variables, e.g., the value of the variable x must within 10 and 100.

Solving such problems may impose high computational overhead, however, with the evolution of *Cloud Computing* and computer power, it became possible to solve large Linear Programs that might consist of millions of variables with thousands of constraints.

1.2 Literature Review

This section provides a critical analysis for the work in the literature that intersects with the addressed problems, and the proposed methodology.

1.2.1 Management in Virtual Private Networks

Previously, till the end of 1990s, the *leased-lines* and *dial-up* telecommunication technologies were almost dominant, Cohen and Kaempfer (2000). Fairly recently, the VPN technology emerged as a new solution that can provide comparable services in terms of reliable, secure, and cost-effective connections. VPNs came with the promise to reduce the expensive costs of the leased-lines technology by avoiding the need of physically leasing the network links from the ISPs. Instead, taking the advantages of the Resource Reservation Protocol - Traffic Engineering (RSVP-TE), Awduche *et al.* (2001), with the Multi Protocol Label Switching (MPLS), Davie and Rekhter (2000) technologies made it possible to provide QoS guarantees over the IP based VPNs.

For QoS provisioning in VPNs, first came the pipe model in Duffield *et al.* (1999b) that attempts buying (leasing) specific bandwidth-pipes between each source-destination couple in the network. Consequently, each node in the VPN is connected by a point-to-point pipe to any other node in the same network, such pipes have certain bandwidth capacities to hold the connections between the considered nodes. This represents an emulation of the traditional frame relay or leased-lines services (like virtual circuits), Duffield *et al.* (2002). Hence, the ISPs would need to continuously provide static bandwidth amounts over each pipe that provides adequate QoS, and SLAs guarantees. Although such a model might provide satisfying service guarantees, on the other hand, it has many limitations, mainly like: (1) It could waste the network bandwidth resources, where resources available at one pipe can not be allocated to another. (2) The VPN customers should have precise knowledge about their *long-term* traffic, expected with the other VPN sites.

Later on, to provide better resource utilization (compared to that in the pipe model), in Duffield *et al.* (1999b), the authors proposed a point-to-cloud model, namely, the hose model. In hose, the model attempts aggregating the incoming/outgoing traffic of one end-point to all other end-points in the network. Thus, by taking the advantage of the statistical multiplexing among the customers' pipes, the network resources can be utilized better. However, although this provides an advantage over the pipe model, in hose, the problem of resource management became more challenging. Indeed, expecting the temporal variations of the hose bandwidth capacities is a challenging task, where there is a big uncertainty of the adequate bandwidth requirement of the VPN end-points. Meeting the customers' SLAs with such weak traffic matrix specifications is not guaranteed. To solve this, in Duffield *et al.* (2002) the authors proposed a model that attempts resizing the hose size in order to provide significant multiplexing gains.

The bandwidth efficiency of the hose model is studied in Juttner *et al.* (2003), where the over provisioning factor of the model is evaluated in networks with various sizes and node densities. The authors conclude that the hose model performs better over the pipe model in reducing blocking probability, decreasing traffic loss, and ease of implementation.

In Gupta *et al.* (2003), the authors gave a randomized approximation algorithm for the general VPN design problem that finds a set of paths between each source/destination couple, where all valid traffic matrices can be routed using these paths. However, their proposed approach has the following limitations:

- the proposed solution is not suitable for VPNs that may hold delay sensitive traffic (e.g. multimedia), as the found set of the paths may not necessarily contain the optimal path;
- the chosen paths are not completely random, in which they may not perform well for real network topologies, where traffic patterns and its desired destinations are changing rapidly.

The authors in Erlebach and Ruegg (2004) proposed a multi-path routing provisioning approach for the hose model. In their work, they ran 6200 series of experiments with small connected random graphs of 3 to 5 nodes. The results indicate that a multi-path routing scheme has a reduced reservation cost compared to that of a shared tree. Roughly the reduction was in average of 20% of the instances with 3 nodes, 25% of the instances with 4 nodes, and 17% of the instances with 5 nodes. In the cases where the multi-path routing had reduced reservation cost compared to tree routing, the average cost reduction was 8.6%.

Accordingly, the authors viewed the results as an indication that multi-path routing has the potential of offering bandwidth savings for VPN reservations in the hose model. However, as discussed earlier, the target of this work is to find an *efficient resource allocation model*. To achieve this target, multi-path routing may not suitable as it may waste the network resources by transmitting duplicated traffic connections in different paths.

In Firestone *et al.* (2007), another work highlights the issue that studies on the hose model deal only with bandwidth requirements, and do not consider providing end-to-end QoS guarantees between VPN endpoints. In fact, this is an important metric especially for those VPNs that carry delay sensitive connection such as multimedia applications.

Consequently, the authors in Ghobadi *et al.* (2007) proposed a ranking approach to enhance the hose model. In their work, they propose guaranteeing the delay requirements between the VPN endpoints while optimizing the provisioning cost. Further, in Ghobadi *et al.* (2008) the same authors presented a new resource provisioning algorithm to enhance the hose model in VPNs. Compared to their prior work, this algorithm provides more efficient results in terms of time complexity, and provisioning cost.

The above studies on resource provisioning and QoS guarantees, of both pipe and hose models, concerned only with finding the optimal provisioning tree, and the cost-effective bandwidth allocations, based on *centralized resource management* mechanisms. With the rapid growth of VPNs users, such centralized resource management schemes are not satisfying anymore, where the need is emerging to reduce the management load at the ISPs side. Virtualized networks, more precisely the *autonomic management*, represents a promising technique that can solve such a problem while providing comparable results in terms of bandwidth utilization, customers' satisfaction, and ISPs profit rates.

1.2.2 Management through the Virtualized Networks

The idea of autonomic resource management have been first introduced in Appleby (2001), where the authors developed a prototype of a highly manageable infrastructure for the e-business computing. Their aim was to develop a manageable and scalable computing infrastructure. Such infrastructures that consists of a farm of parallel packaged servers, interconnected by high-speed switched Local Area Networks (LANs). The concept of *dynamic resource allocation* is mainly developed to accommodate both planned, and unplanned fluctuation of the network state under the constraints of the contracted SLAs.

In HP-Laboratories (2003), and Graupner *et al.* (2003), the Hewlett-Packard (HP) vision for the Adaptive Enterprise and the Microsoft Dynamic Systems initiative Microsoft-Corporation (2004), which are related industry institutions, realize that autonomic management for the computing components is critical for future Information Technology (IT) industry. Moreover, they both gave emphasize to the importance role of the *Virtualiztion* concept of the networks' resources. Consequently, it is clearly noticed that the previous work have mainly focused on the autonomic management of computing resources for IT service delivery.

Works that deployed the autonomic management concept to *network resources* are many. In Lai *et al.* (1998), and Mark *et al.* (2000), the authors proposed a Complete Sharing (CS) model. This CS model proposes an autonomic management framework, where the network customers are authorized to self-manage, and self-control their bandwidth resources. In CS, the network bandwidth resources are all shared between the VPN operators to use, i.e., no virtual partitioning for the bandwidth resources. With no bandwidth divisions, in the CS model, all QoS classes can share the network resources without discrimination. Through this, the model is supposed to ensure the delivery of services according to predefined SLAs. From the ISP prospective, giving the VPN operators such a privilege can provide better utilization of the network resources, assuming that the VPN operators know their changing requirements better, and accordingly they can acquire the required resources based on their actual needs. Naturally, as long as we guarantee that VPN operators will behave *cooperatively*, deploying such an approach can help to reduce the waste of any extra or un-used resources, which results in better satisfaction rates and higher profits at the same time.

However, in reality cooperation between network customers (VPN operators) is not guaranteed, where for many reasons, customers may tend to be non-cooperative and exaggerate their requirements by asking for extra resources. Accordingly, as long as there is no bandwidth divisions, one class may *overwhelm* all other classes. This creates several problems like: SLAs violations, high blocking ratios, and low profit rates.

In Haung and Ho (2002), a Complete Partitioning (CP) model is proposed; another autonomic management model, but with *virtual partitioning*. In the CP model, the scenario is somehow different from that in the CS, in which resources are partitioned among the provided service classes in a *static* way. Hence, it assumes a one-time bandwidth partition for the whole network links. With such a static partitioning scheme, each class is allowed to exclusively use a special portion of the provided resources. This can solve the resource overwhelming problem, but still, it may lead to low bandwidth utilization as resources -of these static portions- might be under-utilized.

A hybrid Virtual Partitioning (VP) approach is proposed by Borst and D.Mitra (1998), this VP runs either as CS or CP depending on the actual network traffic load. Accordingly, it runs as a CS at the light traffic case, while it is a CP and the extreme one. When it runs a CP behavior, VP allows a manual resource sharing between the under-loaded and the over-loaded links (or classes) in order to provide better resource utilization rates. This sounds cool, but the problem here is that the lender links (originally under-loaded) have no guarantees that they can return their resources back when they need it. This creates the problem of QoS violations, and encourages malicious over-loading. Moreover, the resource sharing scheme in VP attempts a static design, in which, a pre-defined static configurations for the resource sharing process is applied at all possible traffic load conditions.

In Farha and Leon-Garcia (2006), and Sup *et al.* (2005), the authors expanded the autonomic view to include the computer telecommunication services. Cheng *et al.* (2006a) proposed an Autonomic Service Architecture (ASA), which is a framework for automated management of both Internet services and their underlying network resources. In this ASA, the authors presented a Bandwidth Borrowing (BR) technique to automate the resource sharing process, and provide a solution for the static load-configuration scheme attempted by the VP approach. While on the contrary, in the resource allocation side, the BR still attempts the same *static resource sharing* scheme. Hence, although the BR scheme can partially solve the bandwidth utilization problem, such solution is not optimal. Consider the case of having many under-loaded links (or classes), with no overloaded ones, in such a case how would the BR help? On the other hand, this *static* partitioning approach -by itself- has several limitations, like:

- deploying a static partitioning scheme means that we are not taking into account the changes of the real-time traffic patterns, and the varying customers' demands. This model will certainly lead to a non-efficient resource partitioning;
- constructing one-time SLAs (based on the static partitioning) is not satisfying for both (ISP, and customers), where market prices are changing dynamically. Increasing market prices can impose profit loses for the ISP, while decreasing prices will reduce the users' satisfaction rates. Competing market prices is a crucial point that is not considered by the CP, VP, and the BR approaches. This issue contradicts with the main objective of maximizing the ISP profits, and the customers satisfaction rates;
- static partitioning motivates the network customers to exaggerate their requirements at the SLAs negotiation phase. Therefore, even if we deploy the BR scheme for better bandwidth utilization, it will not work properly, since it totally depends on what the users reveal. Basically, users may exaggerate their requirements in order to cope with any sudden or unpredictable changes in the network state and link conditions, so they

tend to keep a spare amount of resources that enables them to overcome and cope with such situations.

Although the above studies proposed models that can provide autonomic management schemes, non addressed the anticipated problems and management conflict resulting form such privilege. Rational customers are always eager to maximize their payoffs, unless controlled, cheating is the easiest way to do. In such scenarios, cheating can lead to lower ISP profits, and high blocking ratios. This work proposes new models that can suppress such cheating actions, and motivate the customers to behave truthfully even when they have the privilege of the autonomic management.

1.2.3 Game Theory and Resource Management

Firstly, the concept of Game Theory could be generally defined as a tool that analyzes the conflict and cooperation among intelligent or rational decision makers of a certain system. Thus, it is an excellent tool that could be used in delivering efficient autonomic resource management models that can serve the VNs concept. Such models that rely on the ability to control and adapt the network management strategies in a decentralized manner, without direct interventions from the ISPs. Previous research did not consider the concept of Game Theory in the areas of wired VNs, however, in the domain of wireless networks, Game Theory had been widely applied.

In Etkin *et al.* (2007), and Wang *et al.* (2007), the authors investigated whether spectrum efficiency and fairness can be obtained by modeling the spectrum sharing as a repeated game. Authors in Cao and Zheng (2005) proposed a local bargaining scheme to achieve distributed conflict-free spectrum assignment that can adapt to network topology changes. In Han *et al.* (2007), a no-regret learning algorithm using the correlated equilibrium concept to coordinate the secondary spectrum access was considered.

Various auction and pricing approaches were proposed for efficient spectrum allocation, such as: auction games for interference management, Huang *et al.* (2006), and Gandhi *et al.* (2007), the demand responsive pricing framework, Ileri *et al.* (2005), and pricing for bandwidth sharing between the Worldwide Inter-operability for Microwave Access (WiMAX) networks, and the Wireless Fidelity (WiFi) hot-spots, Niyato and Hossain (2007). A belief assisted distributive double auction was proposed in Ji and Liu (2006) that maximized both primary and secondary users' revenues, and a game-theoretical overview for dynamic spectrum sharing was presented in Ji and Liu (2007).

Although the approaches listed above have boosted the spectrum efficiency, most of them are based on the assumption that the players (e.g., wireless users/devices) are honest and will not cheat. Nevertheless, selfish players aim only to maximize their own interests if they believe that it can gain further increase by cheating. Hence, selfish users will no longer behave honestly, which usually results in poor outcomes from the spectrum sharing game.

Motivated by the preceding, Mechanism Design theory, Ji and Liu (2008), - whose founders: L. Hurwicz, E. S. Maskin, and R. B. Myerson, have recently won the Nobel Prize in Economics in 2007 - emerged as a powerful tool to implement an efficient solution for the decentralized optimization problems, with self-interested players. By carefully setting up the structure of the game, each player has an incentive to behave as the system designer intends, which results the desired outcome.

In this work, the application of Game Theory is also quite possible, as it satisfies the main game components: players, strategy, and utility functions (i.e. VPN operators, operates ideally or selfishly, total profit and QoS guarantees), respectively.

1.2.4 Mechanism Design and Resource Management

Different mechanisms have been proposed in prior research to resolve competitive resource allocation issues for wireless networks in a distributed and scalable manner, Kelly (1997a), Semret *et al.* (2005), and Badia *et al.* (1997). In Kelly (1997a), a pricing mechanism is adopted for resource allocation to ensure that the sum of users' utilities is maximized. However, the users are assumed to be "price takers" (i.e., they do not anticipate the impact of their actions on the network). In Semret *et al.* (2005), it has been shown that resource allocations such as those proposed in Badia *et al.* (1997) suffer from an "efficiency loss" if the users exploit the fact that their actions affect the network prices. In Semret *et al.* (2005), the auction mechanism was deployed for resource allocation. The optimal auction strategies for the resource-buyers are derived and the equilibrium is shown to exist. In Badia *et al.* (1997), pricing schemes are introduced which could be deployed by a service provider to organize the network. However, the relationship between the assigned resources and gained utility is not thoroughly studied.

In Jeffrey *et al.* (1994), authors proposed an approach that claimed to deliver economic efficiency results. They presented a pricing model that achieves incentive-compatible state, where for each served packet, the winner bidder pays a clearing price defined by the highest not-accepted bid (bids of dropped packets). Although this approach might motivate selfish bidders to reveal their truthful information, but on the other hand, it could be classified as not budget-balanced approach, in which, ISPs might provide incentives more than what they actually gain. Authors in Kelly (1997b) proposed an approach that claimed to deliver an efficient pricing model. Their pricing model was built based on a combination of measured and declared traffic characteristics, accordingly an appropriate traffic model is assumed, and corresponding pricing plan is indexed to encourage the bidders to truthfully reveal their information. Although this approach might enhance the resource management process and provide higher profit rates, but it does not address the problem of calculating the optimal prices that bidders have to pay according to the market state. Therefore, one can conclude that this approach does not provide optimal pricing solution.

In Lazar and Semret (1999), a different auction mechanism is proposed to be applied in both additive and arbitrarily divisible resource models. In there work, the proposed model does not consider the QoS as a parameter, where they only assume selling network resources (i.e. bandwidth). This mechanism defines bidders as having an explicit pricing scheme of offered resources, where final selling prices are determined based on the Progressive Second Price (PSP) auction mechanism. Briefly, in PSP, the bidders submit their bids in terms of (required bandwidth and price). Based on the feedback received from the market, bidders reply to any arrival bid by manipulating their offers, either by increasing the price for the same bandwidth amount or keeping the same price but asking for less bandwidth resources. This scenario runs until higher offers raised. Consequently, each winner bidder gains the requested bandwidth amount, and pays the social opportunity cost (second best price). However, such mechanism might deliver non-fair allocations, as it could have more than one truthful-equilibrium point.

The authors of Mohamed *et al.* (2011) have used the mechanism design to motivate the selfish nodes to cooperate and reveal truthfully their private information. Incentives are given to nodes in the form of reputation where the reputation is calculated based on VCG. Nodes grant the services according to their reputation, thus all nodes are motivated to cooperate and truthfully reveal their information in order to increase their reputation value.

Authors of Dramitinos *et al.* (2007) proposed a mechanism that consists of simultaneous Multi-unit Dutch Auctions (MIDAS) for auctioning bandwidth resources over a wide-bases network, to users that will utilize the resources over the same period of time. Links' price-units are assumed to be asymmetric, reflecting the different demands over the various links. In their work, the authors proposed an efficient price reduction policy called the Price Freezing (PF) policy, in which the model deals with the bandwidth pricing problem according to the dynamic bandwidth allocation demands over the considered links or (paths). They assumed that bidders are truth-telling, presenting a payment rule of the VCG type that complements their proposed mechanism, and enforces incentive-compatibility that results in truthful bidding. They also proved that their mechanism provides a promising approach to hard problems, where it requires a low computational complexity and is scalable for large number of units and network users. Never the less, their attained social welfare is close to optimal in general. Moreover, in Nisan and Ronen (1999), it has been used for solving least cost path, and task scheduling problems using algorithmic mechanism design.

Previous works employed the Mechanism Design as an incentive mechanism, while in this work, it is employed as a threat mechanism. Thus, it is motivating the players to reveal the truthful information under the threat of punishment instead of incentives.

1.2.5 Markov Decision Theory and Resource Management

Tran and Dziong (2010) deployed the Markov Decision Process theory (MDP) to propose a decomposed model for bandwidth capacity adaptation in the context of Service Overlay Network (SON). In their work, they intended to maximize the ISPs profit, while maintaining both: the Quality and Grade of service guarantees. The authors depend on the integration of routing and capacity adaptation concepts for maximizing the network profit, and utilizing the network bandwidth resources efficiently. To do so, they followed the real-time measurements fed by special algorithm, based on decentralized dynamic bandwidth allocations.

However, in the case of VPNs, the situation is some how different from that proposed in Tran *et al.* (2007), Tran and Dziong (2010) for SONs. In the case of VPNs, each VPN operator has a limited amount of leased resources (bandwidth over links) to provide the connectivity services within its network according to certain Service Level Agreements (SLAs). On the other hand, from the ISP side, managing the process of resource allocations between different VPNs operators is a more challenging issue. Such a problem requires dealing with different VPNs operators, each with different strategy and utility function. This increases the challenge more and more. In SONs the situation is different, where operators have unlimited pool of resources - changes upon demand - leased directly form the underlying public network (i.e. the Internet), so any demands could be satisfied but with different price.

1.2.6 Linear Programming and Resource Management

In Kumar *et al.* (2001), and Kumar *et al.* (2002), the authors addressed the problem of provisioning in VPNs in the hose model. In their work, they developed algorithms to compute the optimal VPNs' provisioning trees. These trees who require the lowest bandwidth amounts. They assumed different scenarios for the hoses' ingress/egress capacities, i.e., unlimited, equal, and arbitrary. For the unlimited capacity case, they concluded that computing such a tree is NP-hard problem. While in the equal capacity case, they proposed a Breadth-First Search (BFS) algorithm. This algorithm provided the optimal tree with a time complexity of $O(m,n)$, where m and n represents the number of network links and nodes, respectively. However, in the case of having some capacity constraints, the problem became NP-hard, even if the ingree/egress capacities are equal. For the arbitrary capacity case, they propose an integer programming formulation to develop a primal-dual algorithm, but also its computational complexity was way hard. However, compared to the work of Duffield *et al.* (1999b) who employed the Steiner tree, and to the BFS algorithm, the proposed primal-dual algorithm provided the best results represented by the lowest bandwidth requirements.

The author of Chou (2004) considered the problem of TE of MPLS-based VPNs that hold different service classes in a centralized off-line mode. In his work, he focused on considering a multi-objective TE problem that takes into account the resource usage, and links' utilization of Label Switched Paths (LSBs). However, the proposed TE optimization problem found to be NP-complete that involves huge number of variables.

Authors of Juttner *et al.* (2003) presented a comparison between the hose and the pipe models in terms of bandwidth requirements, in the context of the VPNs. Their goal was to compare the requirements of the aforementioned two models with respect to the same network designs, assuming a range of network sizes and topologies. To do so, they proposed deploying a linear programming based evaluation model to calculate the bandwidth usage metrics.

The aforementioned works dealt only with models that do not provide the *autonomic computing* concept. More important, they built their work based on the theme of *static provisioning*. On the contrary, this work addresses the autonomic computing concept in a periodic and dynamic bandwidth provisioning scenarios.

1.3 Summary

This Chapter discussed the tools and theories proposed to achieve the work objectives, along with the related work in the literature. As a summary, Section 1.2 shows that the works done in the field of bandwidth resource management in VPNs are many, however, non provided an optimal solution. Pipe and hose model are the two main approaches proposed for QoS provisioning in VPNs. Still, they did not consider the autonomic management concept, nor the dynamic bandwidth allocation scheme. CS, and CP models deployed the autonomic management, but they have some imitations in terms of QoS violation and low bandwidth utilization. The BR technique is supposed to solve the problem of poor utilization in CP, though, its provided solution is not optimal. Game theory and Mechanism Design have been widely deployed in the areas of wireless networks. Yet, the Mechanism Design has mostly employed to enforce cooperation based on incentives, not as a threat model. The Markov decision theory has been deployed for resource management and capacity adaptation in SONs. In VPNs the scenario is different, where compared to the SONs, the available bandwidth resources is somehow limited. Linear programming theory was also deployed for solving optimal provisioning problems in VPNs, however, neither the autonomic management nor the dynamic bandwidth provisioning were addressed.

CHAPTER 2

A COLLUSION RESISTANT MECHANISM FOR AUTONOMIC RESOURCE MANAGEMENT IN VIRTUAL PRIVATE NETWORKS

Ahmad Nahar Quttoum[1], Hadi Otrok[2], and Zbigniew Dziong[1]

[1]Electrical Engineering Dep., École de Technologie Supérieure, Université du Québec

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2]Computer Engineering Dep., Khalifa University of Science, Technology & Research, UAE

An Article accepted and published by the Elsevier Journal of Computer Communications in August 2010

2.1 Abstract

In this paper, we address the problem of autonomic resource management for Virtual Private Networks (VPNs). Resources management is one of the important problems facing most Internet Service Providers (ISPs). As a solution, the Autonomic Service Architecture (ASA) is proposed in the literature to automate the resources management. Although, this model is able to improve ISPs' performance by automatically adjusting the resource allocation of each customer, it still suffers from two main limitations. First, this model increases the ISPs' revenue in a suboptimal way. Second, this model has no mechanism to prevent customers' *exaggeration* that can lead to a non-efficient resource utilization, and violate the contracted Service Level Agreements' (SLAs) terms. To guarantee their QoS classes; customers might exaggerate by asking for more resources during and after the SLA negotiation session, especially in the case of multimedia streaming, and this can waste the available network resources. To overcome the above limitations, we propose an Autonomic Resources Management Mechanism that increases the ISPs' revenue by allocating resources based on the auction mechanism, where resources are granted to the best bidders. Additionally, we propose a threat model based on Vickrey-Clarke-Groves (VCG) mechanism that is able to penalize exaggerating bidders according to the created inconvenience. Since in our framework, customers are assumed to be

rational, to avoid penalties that represents less profits, they may not ask for more unneeded resources but *collude* with others to have the resources for less. Such a behavior can dramatically minimize the ISP revenue while on the other hand it can maximize the customers' utility. To avoid this, we propose a collusion resistant model based on the Markov Decision theory that allows the ISP to calculate the state dependent optimal cost-unit threshold based on the shadow price concept. All bids that are greater than or equal this threshold are considered in the auction. With this threshold, we can reduce the collusion behavior and with VCG we can motivate the customers to not exaggerate. Simulation results show that the ARMM model is able to efficiently utilize network resources, increase ISPs' profit, and customers' satisfaction rates.

2.2 Introduction

Virtual Private Networks (VPNs) use the infrastructure of Internet Service Providers (ISPs) to establish secure and reliable streams of services, Duffield *et al.* (1999a). ISPs are in need for a flexible and efficient management model that is able to support a wide variety of customers and satisfy their needs, in terms of secure and reliable connections with competitive prices. Relying on the current management models that attempt direct interactions with ISPs is not satisfying anymore, as they have many limitations that could be summarized as follows: (1) Such models increase the management operation expenses. (2) They provide slow response times. These limitations can lead to high rates of customers' dissatisfaction. Consequently, the need is emerging to find an alternative resource management models that overcome the current limitations. Autonomic Service Architecture (ASA) is proposed in Cheng *et al.* (2006a) to cope with the above limitations by creating a uniform framework for automated management. ASA ensures services' delivery based on a Service level Agreements (SLAs) that have been conducted between customers and ISPs. The aim of the model is to propose an efficient resource management scheme that can increase the revenue of ISPs while maintaining satisfactory QoS guarantees. Although the ASA is able to automatically adjust the resource allocation of the customers through deploying an autonomic bandwidth borrowing scheme, still it has the following limitations:

- it increases the ISP's revenue in a suboptimal way by allocating resources based on the First Ask First Allocate (FAFA) concept. This is due to the fact that resources can be allocated to the non-optimal set of customers, and therefore ISPs' profit will not be maximized;
- customers (i.e. VPN operators) can exaggerate and ask for more resources than needed during the SLA negotiation phase to guarantee their QoS in order to cope with the unpredicted variations in the network state, especially in the case of multimedia transmissions. Such a behavior can lead to an unfair resource allocation, and significantly decrease the ISP's revenue.

In this paper, we consider deploying an *Auction Mechanism* allocating the network bandwidth resources among VPN customers. Although the auction mechanism optimizes the ISP's profit, it is unable to solve the problem of exaggeration. Therefore, we are also proposing a model that governs revealing the truthful requirements under the threat of punishment. In this model, we deploy the well known *Vickery - Clark - Groves* (VCG) truth-telling mechanism, Nisan and Ronen (1999), to calculate the inconvenience that each VPN operator causes to the whole network according to its required resources. This inconvenience is defined in terms of the utility drop that is caused to the whole network. The resulting value is denoted by the "*transfer*", note that this transfer value is always negative or zero representing the occurred inconvenience. Hence, the transfer value will be added to the original charge of the leased resources, which may result in smaller utility values for the exaggerating operators. Consequently, VPN operators will not ask for more resources than their real needs, as they know that exaggeration will decrease their utility. Collusion is another possible strategy that bidders can follow to maximize their utility by having the resources for less. This strategy can dramatically minimize the ISP revenue. To maximize the ISP profit and reduce collusion, an optimal cost-unit threshold is needed. We propose a Markov Decision Process (MDP) based approach to calculate a state-dependent selling cost-unit threshold, where bids that are greater than or equal the threshold are accepted in the auction. More specifically, we employ the concept of *shadow price* in order to help the ISP to determine the optimal cost-unit at every state. By calculating such a threshold, we reduce collusion and by having VCG we reduce the exaggeration behavior.

In summary, our contribution is a model that is able to:

- efficiently utilize the available network bandwidth resources, since customers are motivated to reveal their truthful needs;
- suppress exaggeration actions via a threat model;
- increase the customers' satisfaction rates since resources are utilized efficiently;
- maximize the ISPs' revenue by reducing the effect of collusion.

2.2.1 Paper Organization

The rest of this paper is organized as follows: Section 2.3 presents the problem statement. Section 2.4 presents our resource allocation model and ARMM selection algorithm. Sections 2.5 and 2.6 respectively illustrate the ARMM threat and collusion-resistant models followed by an illustrative example in Section 2.7. Section 2.8 presents empirical results. In Section 2.9, we present the related work. Finally, Section 2.10 concludes the paper.

2.3 Problem Statement

The ASA model proposes an autonomic management framework that is able to ensure the delivery of services according to predefined SLAs. These SLAs are established after negotiation between an ISP's broker and customers. The objective of the ASA is to increase the ISP's revenue by providing an efficient resource management scheme. ASA is considered as a SLA central management model that assumes the share of resources among all SLAs. To achieve this, the model proposes an autonomic bandwidth borrowing scheme for efficient resource utilization to ensure customers' QoS.

The ASA has several limitations. First, resources in ASA are utilized through an inefficient way by providing the resources in a First Ask First Allocate (FAFA) scheme. This model can lead to a suboptimal increase in the ISP's revenue. Second, the negotiation phase and the bandwidth borrowing scheme depend totally on what the customers reveal/ask, where customers might exaggerate with their revealed requirements and thus some network resources can be wasted. Customers exaggerate for different reasons such as in the case of multimedia transmission, where:

a. the quality of the transmitting streams increases with the higher transmission rates, and the more bandwidth resources allocated. Without having any threat mechanism, customers will always try to obtain the largest possible amount of bandwidth resources over network links, even if the resulting improvement in the transmitting quality is minimal;

b. in order to cope with any sudden and unpredictable changes in the network state and link conditions, VPN operators tend to keep a spare amount of resources that enables them to overcome and cope with such situations. Again, as there is no threat scheme, VPN operators will tend to obtain as much as they can of bandwidth resources which can block others;

c. if VPN operators are allowed to obtain more resources than required, they have no incentives to smartly utilize their allocated resources and use it efficiently. Instead, they may rely on their ability to resend lost and delayed packets. This achieves a relatively good transmission qualities while using their traditional transmission techniques.

To overcome the above limitations, we propose to allocate resources based on an auction manner. This can improve resource utilization and increase the ISP's revenue. To overcome the exaggeration problem, we propose a new mechanism that urges VPN operators to truthfully reveal their requirements, and respect the SLAs' terms. This mechanism adopts the well known VCG truth-telling mechanism that is able to handle such a problem by motivating VPN operators to truthfully reveal their requirements and respect the SLAs' terms under the threat of punishment. The threat is expressed in terms of a "*transfer*" value that is derived in Section 2.5.

Still, if exaggeration actions are controlled under the threat of punishment, VPN operators may search for another behavior that can help to increase their utilities. One possible behavior can be the collusion among the bidding operators, where VPN operators can act collusively against the ISP in order to allocate the resources according to the price they revealed. More precisely, a group of VPN operators may collude and offer the same low bid. According to the bid rules, an ISP has no other choice than allocating the resources to the colluded bidders. Such a collusive cheating behavior among system players poses serious threats to the ISPs' revenue, and to the whole resource allocation game outcomes. To reduce the effect of this problem, we approximate the auction game by a *Markov Process* describing the system states by the aggregated bandwidth requirements at each auction slot. Then using the *shadow price* concept based on Markov decision theory, we define the optimal allocation cost-unit to be considered as a threshold for accepting or not-accepting the received bids, where this threshold is not known to the bidders. The model for calculating such *shadow price* values is presented in Section 2.6.

2.4 ARMM: Autonomic Resource Management Mechanism

In this section, we present the proposed ARMM model that improves the ASA model by allocating resources based on the auction mechanism, where resources are allocated to the best bidders. To achieve this, bidders are asked to reveal their required QoS classes and their respective prices. The revealed QoS information represent the required bandwidth resources to be allocated to each bidder. An auction algorithm is presented in Algorithm 2.2 to illustrate the allocation mechanism. Although auction mechanism can increase ISPs' revenue by creating a competition environment among the bidders, but it cannot prevent bidders' exaggeration nor collusion. To overcome the exaggeration problem, we propose a *threat model* that can punish bidders according to the inconvenience they create to others. This inconvenience is calculated based on the VCG mechanism. To reduce the collusion problem, we derive the optimal state dependent *threshold* for accepting the bids during the auction process. This threshold is calculated based on the shadow price concept. Tables 2.5 and 2.6 in section 2.7 provide illustrative examples to show how ARMM is able to penalize exaggerated bidders and reduce collusion actions, respectively.

2.4.1 The Model

The resource allocation problem can be modeled as a game where the VPN operators are the players of the game. The players are assumed to be *rational*, and thus their aim is to maximize their own *utilities* according to the revealed values of their required *QoS* and offered *Prices*. The offered price value implicitly represents the anticipated *utility-gain* (ρ) that the player can collect from this connection (the term connection refers to the allocated resources), and always aims to maximize. Utility-gain maximization leads to higher utility rates, where the player's utility is represented as the aggregation of its utility-gain (ρ) and the system's *transfer* value (τ). This function is expressed as follows:

$$Utility_i = utility.gain_i + transfer_i \tag{2.1}$$

On the other hand, the objective of the system is to maximize the sum of bids while maintaining the QoS guarantees for the whole network.

2.4.2 Bandwidth and Cost Measurement Model

The VPN operators have different service classes to choose their required connectivity QoS levels. First, VPN operators submit their bids (price, QoS) to an ISP's broker. Second, the ISP broker calculates the required bandwidth allocations based on the information received. ITU-T recommends to collect the operators' QoS judgement classes using different scale levels, ITUT-Rec-P800 (2000), as shown in Table 2.1.

Consequently, we adopted the ITU-T model to define the ISPs' provided QoS classes. Such class factors could be converted to the Mean Opinion Score (MOS) that provides a numerical interpretation of the required connections' quality ITUT-Rec-P800 (2000). Thus, with the operators' MOS values, the ISP broker can determine the codec required for encoding/decoding the communicating packet streams. The bandwidth amount required by a VPN operator's connection is typically defined according to the size of headers and packet payloads, and the type of codecs. Therefore, the following formulas are used to calculate the bandwidth consumption per allocation request:

Table 2.1 QoS Classes

Class	**Connections' Quality of Service**
5	Voice and Video ($<$100ms Latency and Jitter)
4	Controlled Load (Streaming Multimedia)
3	Excellent Load (Business Critical)
2	Standard (IP Packet Delivery)
1	Best Effort

- total packet size = headers + packet payload size;
- packets per second = codec bit rate / packet payload size;
- bandwidth = total packet size * packets per second.

Thus, operators reveal their (price, QoS) values, and then the ISP broker performs the mentioned calculations to determine the requirements of each VPN operator, and consequently check the offered price if it is accepted or not as explained in the following subsection.

Based on the offered and required values of the VPN operators (prices, bandwidth), our model measures the *cost-unit* based on the ratio of the offered prices to the required bandwidth amounts, as follows:

$$c_i = \frac{price_i}{bw_i} \tag{2.2}$$

where:

- c_i: represents the measured cost-unit;
- $price_i$: represents the price offered by VPN operator i;
- bw_i: represents the the bandwidth needed for operator i to satisfy its required QoS class.

2.4.3 The ARMM Selection Algorithm

Based on the measured cost-units, the proposed model will select the VPN operators with the most profitable *cost-units* that suppose to maximize the ISP's profit, in accordance to the contracted SLAs and bandwidth constraints. Algorithm 2.2 describes the selection algorithm.

Algorithm 2.2 Selection Algorithm

Selecting the Profitable VPNs Operators

1: VPN operator i submit ($price_i$, QoS_i)$\rightarrow$ ISP broker;
2: **for** each VPN operator i, **do**;
3: Convert QoS_i class to its equivalent bw_i value;
4: Calculate c_i for each VPN operator i;
5: Sort the VPN operators in a descending order according to their c_i values ;
6: Find i_{opt}, where;
$\sum_{i=1}^{i_{opt}} (bw_i \leq bw)$ and $\sum_{i=1}^{i_{opt}+1} (bw_i > bw)$
7: **Output** (i_{opt}) operators that fit within the available bw;
8: Charge the selected (i_{opt}) operators according to their c_i;

In the first step, VPN operators are asked to submit their offered price and QoS values to an ISP broker in order to consider their requests. In step two, the ISP broker takes each VPN operator (price and QoS values), and then computes the required *bandwidth* amount and the *cost-unit*, in steps three and four respectively. In the fifth step, the ISP broker sorts the VPN operators according to their offered cost-units in a descending order. Then in steps six and seven, it checks the available bandwidth amount and accordingly chose the optimal VPN operators i_{opt} according to their offered cost-unit values, where i_{opt} is the index of the last VPN operator (sorted in a descending order) that can fit within the available bandwidth resources. In step eight, the ISP broker charges the selected i_{opt} operators according to their costs. The charge is calculated in the sequel section.

2.5 The ARMM Threat Model

In the current management models such as ASA, the absence of any incentive mechanism that motivates VPN operators to reveal truthfully their requirements can motivate the operators to exaggerate in their requirements if such a behavior can lead to even a minimal maximize in their own utilities. Particularly, in congested networks, if a VPN operator lies or exaggerates about its bandwidth requirements, the performance of the entire network can be effected negatively. Hence, there is a need to develop a model that guarantees the integrity of bandwidth resources and protect it from being wasted or misused. For this reason, we propose a model that overcomes any exaggeration behavior and motivate VPN operators to truthfully reveal their bandwidth requirements. As mentioned before, we adopted a mechanism from the subfield of game theory which is known as the VCG mechanism, Nisan and Ronen (1999), to calculate the *transfer* value, by which ISPs can enforce the VPN operators to cooperate under the threat of punishment based on the inconvenience each VPN operator causes to the whole network.

To address the above challenge, in ARMM, we propose using the VCG truth-telling mechanism to define an "*optimal decision*" $T(\theta, G)$ that provides fair allocations of network bandwidth resources among VPN operators, and also, defines the "*transfer*" value $\tau(\theta, G)$ that represents the inconvenience each VPN operator will cause to the other competing operators. Where θ represents the "profile type" for the VPN operators which includes: 1) The utility-gain per unit time, and 2) The connection's QoS class, while G represents the total amount of resources available at the ISP's premises. The transfer value should be added to the utility-gain value, together giving the total *Utility* value υ of this new VPN operator (i). Using the above notations, Equation 2.1 can be rewritten as follows:

$$\upsilon_i = \rho_i + \tau_i \tag{2.3}$$

Obviously, the objective of each VPN operator is to maximize its *Utility* function. On the other hand, the objective of the ISP is to maximize whole system's utility function which is:

$$U^{sys}(T(\theta,G),\theta) = \sum_{i=1}^{i_{opt}} \rho_i(t_i,\theta_i) \tag{2.4}$$

where t_i represents the amount of bandwidth allocated to VPN operator i according to the value of θ_i. Consequently, the *optimal decision* should allocate the networks' bandwidth resources G among the competing VPN operators in a way that maximizes the aggregated utilities of all VPN operators. Thus, the *optimal decision* is defined as follows:

$$T^{opt}(\theta,G) = arg\ max\ U^{sys}(T(\theta,G),\theta) \tag{2.5}$$

Hence, the transfer function for VPN operator i is computed as follows:

$$\tau_i(\theta,G) = \sum_{n\neq i} \rho_n(t_n{}^{opt},\theta_n) - {}_{T_{-i}(\theta_{-i},G)}^{MAX} \sum_{n\neq i} \rho_n(t_n,\theta_n) \tag{2.6}$$

where, θ_{-i} indicates the profile type of all VPN operators except operator i, i.e., θ_1, θ_2, θ_3, ..., θ_{-i}, θ_{+i}, ..., θ_M, and $T_{-i}(\theta_{-i},G))$ indicates the bandwidth allocations for all VPN operators except operator i, i.e., t_1, t_2, t_3, ..., t_{i-1}, t_{i+1}, ..., t_M.

In this Equation, the first term represents the sum of the aggregated utilities -in the presence of operator i- of all other VPN operators given by the optimal bandwidth allocations except VPN operator i under non-optimal allocation. The second term represents the maximum sum of aggregated utilities that all VPN operators can obtain if VPN operator i does not participate in the bandwidth allocation game. Clearly, the second term will be always greater than or equal the first term, this means that the "*transfer*" value will always be negative or zero representing the inconvenience (utility drop) caused to other VPN operators by operator i. Accordingly, Equation 2.3 that gives the total utility value can be reformulated as:

$$\begin{aligned}
\upsilon_i(\theta_i, t_i) &= \rho_i + \tau_i \\
&= \rho_i + \sum_{n \neq i} \rho_n({t_n}^{opt}, \theta_n) \\
&- \underset{T_{-i}(\theta_{-i}, G)}{MAX} \sum_{n \neq i} \rho_n(t_n, \theta_n) \\
&= [\rho_i + \sum_{n \neq i} \rho_n({t_n}^{opt}, \theta_n)] \\
&- \underset{T_{-i}(\theta_{-i}, G)}{MAX} \sum_{n \neq i} \rho_n(t_n, \theta_n)
\end{aligned} \tag{2.7}$$

By using Equation 2.7, VPN operators will not be motivated to exaggerate their bandwidth requirements, but instead they will tend to reveal their truthful requirements in order to avoid any extra expenses and charges paid according to the required amounts of bandwidth resources, which may reduce their final revenue rates. However, it is worth to find the computational complexity of the model. First, all bids have to be sorted in a descending order where the complexity of this is $O(n.log(n))$. Then, best summation of bids is calculated where the complexity is $O(m)$, where m is the number of selected bids which is assumed to be less than n in order to have a competition game. This will be repeated for m bidders so the complexity will be O(m^2). Thus the total computational complexity will be $O(n.log(n))$ + $O(m^2)$ which can be approximated to $O(n.log(n))$ where $n >> m$.

Although mechanism design can motivate the bidders to bid truthfully their requirements, however, rational VPN operators will keep looking for another behavior to maximize their utilities. Collusion is one of the possible behaviors that may significantly effect the anticipated objectives of the resource allocation game. More details about our proposed model for solving this problem are given in section 2.6.

2.6 The ARMM Collusion-Resistance Model

In our framework, customers (VPN operators) are assumed to be rational, and therefore they may tend to cheat whenever they believe that such a behavior can provide even a minimal increase to their utilities. In the proposed auction model, exaggeration actions can be suppressed by the threat model, but still, VPN operators might collude and bid the same value with the incentive of obtaining lower allocation prices that maximize their utility functions. Such a collusive action creates a great threat to the anticipated efficiency of the proposed resource allocation model. Hence, the need is emerging to develop a collusion-resistant model that can reduce the effect of such collusion actions, in order to provide a better resource allocation model. In this section we propose such a model that is based on optimal cost-unit threshold derived from the Markov decision theory.

2.6.1 Optimal Cost-Unit Threshold

We assume that for a given VPN, the links connecting its nodes are established by leasing bandwidth from a public network infrastructure (i.e the Internet) through SLAs. For each link, the SLA terms specify its assigned bandwidth, its QoS class, and its corresponding leasing cost unit C_o. Here, we assume that the SLA terms can be dynamically modified according to the main provider's network state.

Consequently, from the ISP perspective, the process of sub-leasing the available resources to the VPN operators must be controlled in a way that provides profit maximization guarantees and satisfactory services. Relying on the proposed *auction mechanism* for selecting the best offered prices may enhance both profit and satisfaction rates for all (ISP and VPN operators).

Deploying the VCG mechanism also provides an efficient *threat model* that eliminates exaggeration actions. However, neither the auction mechanism nor the threat model can solve the problem of collusion. Therefore, in this section, we are enhancing the proposed auction mechanism by developing a model that promises to reduce the effect of the collusion problem by *dynamically* calculating the optimal sub-leasing cost-unit threshold, which will be considered as a threshold for accepting/not-accepting the received allocation bids. Calculating this cost-unit threshold is a function of the reward parameter, r, anticipated from each accepted connection. The ISP's utility objective is given by the following:

$$Utility^{ISP} = Reward - Cost \tag{2.8}$$

where the ISP utility function is represented by: the rewards collected from allocation requests being granted, subtracted by the original cost of the system's bandwidth resources.

In the considered model, the i^{th} bidder is characterized by the following: offered cost-unit c_i, and required bandwidth amount bw_i. We assume that at each periodic auction slot, each admitted VPN connection i brings to the ISP a reward given by the parameter r_i. Accordingly, the utility (profit) that an ISP can collect at an auction slot in which the system is in state l, can be expressed as follows:

$$U_l^{ISP} = R_l(bw) - C_o(bw) = \sum_{i=1}^{i_{opt}} \bar{\lambda}_i(bw_i)r_i - C_o(bw) \tag{2.9}$$

where U_l^{ISP} represents the utility of the ISP, R_l is the link reward rate,$\bar{\lambda}_i$ is the rate of admitted allocation requests, and C_o is the original leasing cost unit (the cost paid by the ISP for leasing the bandwidth resources from its higher provider). In this context, we assume that the key objectives for the ISP are: 1) Reduce the effect of collusion. 2) Provide better utilization of the available bandwidth resources. 3) Maintain profit maximization; all with respect to the sub-leasing cost units c_i, being a function of the allocated bandwidth resources bw. This scenario could be considered as a classical bandwidth allocation problem formulation, however, in this work we are solving such a problem by adapting the optimal allocation cost-unit threshold c_{thr} that is a function of the link state.

For cost-unit threshold calculations, we propose to use the notation of *state-dependent shadow price* p_l that represents the dynamic cost of admitting the allocation requests for each bandwidth unit bw_j, where $\sum_j bw_j = bw$, on the considered link in state l, $(l \in X)$. The notion of state-dependent shadow price is derived from the Markov decision theory and will be presented in the next subsection including model for its calculation.

As already mentioned, once the shadow price is calculated, it is considered as the optimal cost-unit threshold c_{thr} for each bandwidth unit bw_j. Therefore, the model assigns higher subleasing costs (rewards required for granting the allocation) to those allocations whose bandwidth requirements are costlier to the network.

Accordingly, the allocation criteria presented in Algorithm 2.2 is enhanced in Algorithm 2.3, where in step 6, the process of accepting or not-accepting the received bids will be through comparing the submitted cost-unit c_i values with the cost-unit threshold c_{thr}, calculated using the *state-dependent shadow price* concept.

To illustrate the procedure, in Figure 2.1 we give an example of the offers received from the VPN operators in terms of their submitted bids sorted in a descending order, and the measured cost-unit threshold for each link state. The process of finding the lowest accepted bid value is to find the crossing point between the resulting optimal cost-unit threshold (c_{thr}) and the current bids sorted in a descending order. Therefore, by using the optimal cost-unit threshold, the ISP is able to reduce the effect of collusion actions in the sense that, no bid will be accepted unless $c_i \geq c_{thr}$, as:

Algorithm 2.3 Selection Algorithm 2

Selecting the Profitable VPNs Operators for each bw$_j$

1: For each auction round (new state), find c_{thr} for each bandwidth unit bw$_j$ over the assigned link;
2: VPN's operator i submit ($Price_i$, QoS_i)$\rightarrow$ ISP broker;
3: **for** each VPN operator i, **do**;
4: Convert QoS$_i$ class to its equivalent bw$_i$ value;
5: Calculate c$_i$ / bw$_j$ provided by each VPN operator i;
6: if c$_i \geq c_{thr}$, then accept for competition, else reject;
7: Sort the accepted VPN operators in a descending order according to their c$_i$ values ;
8: Select the highest c$_i$
9: end;
10: **Output** the selected operator i that won the allocation of bw$_j$;
11: Charge the selected (i^{th}) operator according to its c_i;

$$c_1 \geq c_2 \geq \ldots \geq c_{thr-1} \geq c_{thr} \geq c_{thr+1}$$

Consequently, the value c_{thr} acts as a successive threshold, where the lowest accepted bid must exceed it. As an example, if we look at state 2 in Figure 2.1, we can clearly notice that all the bids are accepted for competition as they are all higher than the measured cost-unit threshold, where in state 3 two bids only (bids 1 and 2) are accepted as they are higher than the measured threshold.

In this way, ISPs can minimize the effect of collusion actions on their utility functions and revenue objectives. In the case where the ISP did not receive any higher bids that satisfy the threshold constraints, it has no interest to lease any more resources as it is assumed that the resources are leased from a third party which is the infrastructure provider.

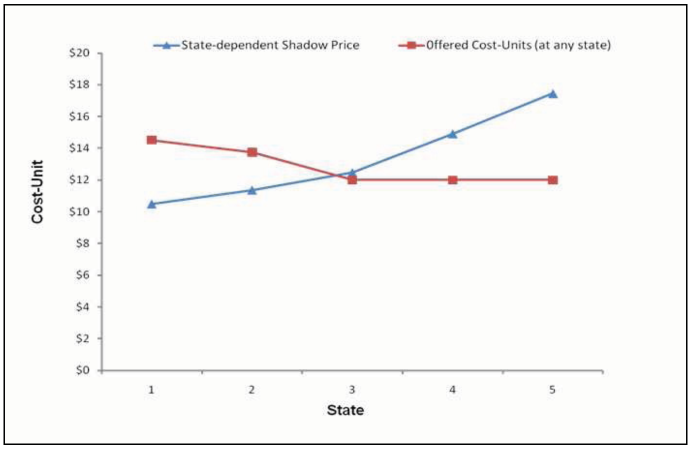


Figure 2.1 Cost-unit thresholds and offered bids

2.6.2 Resource Allocation based on Markov Decision Process

The resource allocation game proposed in the previous sections represents a dynamic system whose evolution is described by a sequence of system states $\{x_t, t = 0, 1, ...\}$ belonging to a discrete state space X. The state here is described by the bandwidth amounts required to satisfy the allocation demands at each time instant t. The time t corresponds to the system review instances representing the auction slots that are held in equal time intervals. When the system arrives at time t, it is classified into state $l = x_t \in X$. Based on the system state, an action is made defining the winning bids. Each action results in a certain reward, R,which is given to the system once the decision is executed during the slot associated with the new state. Accordingly, in the next time epoch, the system moves to the next state based on a new action. It should be noted that the reward given to the system at the decision moment, can also represent the expected reward collected until the next decision moment.

A prescription for the final decision in each time instance is called policy π. To find an optimal policy for our system we use the concept of shadow price that was developed for Call Admission Control (CAC) and routing in telecommunication networks, which can be modeled as a continuous time Markov process Dziong (1997). In this case the network is treated as Markov Decision Process with infinite planning horizon.

To simplify the problem, the network process is decomposed into a separable link Markov processes that are assumed independent. In this case the optimal CAC and routing decision is found by policy iteration algorithm that is based on the concept of net-gain, which is the difference between the reward from the connection and the sum of state-dependent link shadow prices over the links constituting the considered path. In this case, the path with maximum positive net-gain is selected and if it is negative the connection is rejected. Note that in the system considered in this paper each link auction is considered independently and therefore the problem is reduced to a link problem. In such a case it is easy to show that the optimal acceptance policy is defined by the values of link shadow prices. In particular, the connection is accepted in given link state if the reward from connection is bigger than the shadow price for this state.

As shown in Dziong (1997), the state dependent link shadow price can be calculated using efficient value iteration algorithm. Moreover, in homogeneous bandwidth requirement cases, a simple recurrence (over the states) solution can be used. In this paper we apply this model to approximate the state-dependent prices in our system. For this case the calculation of state dependent shadow prices is as follows:

First the values of the recurrences $z(j)$, and $w(j)$ are calculated as follows using the initial values of $z(0) = 1/\lambda(0,\pi)$, and $w(0) = 0$:

$$z(j) = \frac{1 + j\mu z(j-1)}{\lambda(j,\pi)}, j = 1, ..., N-1 \tag{2.10}$$

$$w(j) = \frac{j\mu w(j-1) - q(j)}{\lambda(j,\pi)}, j = 1, ..., N-1 \tag{2.11}$$

where N denotes the maximum number of allocations (available bandwidth-units) that can be held over the link with the available bandwidth bw, and $q(j)$ is the rate of reward defined as $q(j) = j\mu.r$ where r is an average reward parameter.

Then, the state-dependent link net-gains $g_j(\pi)$ and the average reward rate, $\bar{R}$, can be calculated as follows:

$$g_j(\pi) = \bar{R}(\pi)z(j) + w(j), j = 0, ..., N-1 \tag{2.12}$$

the term g_j refers to the network gain at state l, $l \in X$. The average reward $\bar{R}(\pi)$ is given by:

$$\bar{R}(\pi) = \frac{q(N) - N\mu w(N-1)}{1 + N\mu z(N-1)} \tag{2.13}$$

Having the network state gain values g_j for each state, we can find the state-dependent shadow price value p_j, for each bw_j, as follows, [Dziong (1997)]:

$$p_j(\pi) = r - g_j(\pi) \tag{2.14}$$

It is worth to mention that the notion of average link shadow price was introduced first by Kelly (1988) for optimization of adaptive load sharing and in Dziong *et al.* (1988) the authors introduced the state-dependent link shadow price notion that was used for optimization of state-dependent routing.

2.7 Illustrative Example

In this section, we present an illustrative example that shows the computation of the transfer and threshold values.

2.7.1 Impact of adopting the VCG truth-telling Mechanism

To assess the efficiency of the proposed *transfer* function in Equation 2.6 on the resource allocation game, we illustrate an example of five different VPN operators competing for 600 Mbs of available network bandwidth resources at the ISPs' premises. For the VPN operators, we compare the incurred transfer value of each VPN operator and the resulting revenues in terms of the connections' utility functions under two different scenarios: 1) All VPN operators deploy their optimal strategies, and no operator is exaggerating its revealed type. 2) VPN 3 operator is exaggerating its type by asking for higher QoS class (more bandwidth allocation), while the rest operators are revealing their truthful requirements.

The predefined QoS classes and their minimum accepted cost-unit parameters of the deployed example are summarized in Table 2.4. Table 2.5 is showing the percentage of *bandwidth resources* allocated to the various competing VPN operators, their required *QoS* classes, their anticipated *utility-gains* (represented by the revealed price values), and the corresponding *transfer* values for both scenarios. To improve the readability of the results, the difference in the resulting *utility* values between the two scenarios is also given based on Equation 2.7.

Table 2.4 Connections' classes and corresponding rates

QoS Class	Minimum cost-unit accepted
1.0 - 1.5	1.3 / time unit
1.6 - 1.9	1.8 / time unit
2.0 - 2.5	2.3 / time unit
2.6 - 2.9	2.8 / time unit
3.0 - 3.5	3.3 / time unit
3.6 - 3.9	3.8 / time unit
4.0 - 4.5	4.3 / time unit
4.6 - 5.0	4.8 / time unit

In Table 2.5, when VPN operators adopt their best strategies and reveal their truthful bandwidth requirements, network resources are allocated in a manner that maximizes the whole network's utility providing an optimal allocations for all. However, when VPN operator 3 exaggerate, the provided QoS is improved to class 3. On the other hand, the corresponding transfer value for this operator is also increased from 10% to approximately 23% of the revealed utility-gain value. Hence, operator 3 earns QoS class 3, but paid around 9.5 unit more compared with the regular charges, where instead of paying 7.326 it paid 16.785.

Table 2.5 Resource Allocation, Transfer and QoS for various VPN operators in two scenarios (Scenario A: No VPN's Operator Exaggerate, B: VPN 3 Operator is Exaggerating and reveal selfish requirements)

VPN's Operator	*A: No VPN Operator Exaggerate*					*B: VPN 3 Operator is Exaggerating*					*Revenue gained in Scenario A*	*Revenue gained in Scenario B*
	% of Bandwidth Resources	*QoS Class*	*Cost Unit*	*Revealed Utility-gain*	*Transfer Value*	*% of Bandwidth Resources*	*QoS Class*	*Cost Unit*	*Revealed Utility-gain*	*Transfer Value*		
1	11.10	1.13	1.3	14.43	-1.443	10.60	1.00	1.3	13.78	-1.378	12.987	12.402
2	15.10	1.65	1.8	27.18	-2.718	14.60	1.51	1.3	18.98	-1.898	24.462	17.082
3	17.20	2.00	2.3	39.56	-3.956	**22.20**	**3.00**	**3.3**	**7.326**	**-16.785**	**35.604**	**56.475**
4	25.20	3.70	3.8	95.76	-9.576	23.20	3.20	3.3	76.56	-7.656	86.184	68.904
5	31.40	5.00	4.8	150.72	-15.072	29.40	4.50	4.3	126.42	-12.642	135.648	113.778

The example also conclude that the exaggeration of operator 3 affects the performance of the whole network customers, leading to a reduction in their provided QoS classes along with a hard decrease in their revenue values. From this example, it is clear that applying the VCG truth-telling mechanism is significantly decreasing the tendency of exaggeration by VPN operators since VPN operators are rational and they care about their revenue. Thus, under the threat of paying high transfer values, VPN operators will tend to reveal their truthful bandwidth requirements and not to ask for extra un-needed resources. Moreover, this example also shows that relying on the ASA model may result in a significantly worse performance, less resource utilization, and lower profit rates especially when VPN operators start exaggerating their requirements.

2.7.2 Impact of adopting the Collusion-Resistance Mechanism

To assess the efficiency of the *collusion-resistance* model proposed in Section 2.6, another example is presented that illustrates the collusive behavior for a group of different VPN operators that collude instead of competing with each others to be granted a connection over a single link that can hold five bandwidth units. For this link, we compare the resulting reward earned by an ISP in terms of the subleasing cost-units defined by the two scenarios: 1) According to the highest bids received (Selection based on Algorithm 2.2). 2) According to the measured

c_{thr} value based on the state-dependent shadow price concept, that is dynamically measured at every new state $l \in X$ (Selection based on Algorithm 2.3). Through this, we show how the proposed collusion-resistance mechanism is able to reduce the effect of such collusive actions by dynamically adapting the subleasing cost-unit threshold according to the network state and the SLAs cost variations. This example shows the behavior in which a group of VPN operators collude with each others, where they bid with the same bid in order to enforce the ISP to offer them the connections for less, with the incentive of maximizing their own utilities.

Table 2.6 Bandwidth Units Allocation, Offered and Eligible prices for various VPN operators in two scenarios (Scenario A: No Collusion Resistance Mechanism, B: VPN 3 Collusion Resistance Mechanism Using the Measured, c_{thr}, values)

Bandwidth Unit Index	*No Collusion-Resistance Mechanism Applied*			*Collusion-Resistance Mechanism is Applied*				*Aggregated reward in Scenario A*	*Aggregated reward in Scenario B*
	Highest Offered Cost-Unit	*Accept or Reject*	*Eligible Cost-Unit*	*Highest Offered Cost-Unit*	*State-Dependent Shadow Price (c_{thr})*	*Accept or Reject*	*Eligible Cost-Unit*		
1	14.520	*Accept*	14.520	14.520	10.494	*Accept*	14.520	14.520	14.520
2	13.740	*Accept*	13.740	13.740	11.374	*Accept*	13.740	28.260	28.260
3	**12.000**	*Accept*	12.000	12.000	12.467	*Reject*	**12.467**	40.260	40.727
4	**12.000**	*Accept*	12.000	12.000	14.092	*Reject*	**14.092**	52.260	54.819
5	**12.000**	*Accept*	12.000	12.000	17.453	*Reject*	**17.453**	**64.260**	**72.272**

Table 2.6 is showing the link's bandwidth units, bw_j, that are offered to the VPN operators to carry their connections, the highest bid (offered cost-unit) received for the assigned bandwidth-unit, the decision that an ISP will take (accept/reject) depending on the adopted allocation mechanism, the state-dependent shadow price value for each bw_j (this is in the collusion-resistance scenario only), and the eligible subleasing cost-unit. To improve the readability of the results, the difference in the resulting aggregated reward values (anticipated value in scenario B) between the two scenarios is also given.

The example in Table 2.6 shows the case where a group of VPN operators offered the same collusive bid (cost-unit) with the value 12 (for bandwidth units 3, 4, and 5), and this value was the maximum value received. In the case of no collusion-resistance mechanism applied, as long as there is enough bandwidth resources available, the ISP will accept the offered bids based to the selection criteria presented in Algorithm 2.2. Accordingly, such collusive bids will be eligible, and therefore the resulting reward will be less. While in the second scenario, the state-dependent shadow price value, c_{thr}, is computed for each bandwidth-unit producing an acceptance condition that acts as a collusion-resistance threshold. In this case, bids below the cost-unit threshold, c_{thr}, will not be accepted, and so, as long as VPN operators know that their collusive bids will not go throw, they will not tend to collude. Moreover, in the case of no higher bids received (higher than the computed threshold), ISPs can return their extra (un-leased) resources to their higher providers instead of sub-leasing them with such non-satisfactory prices. Additionally, applying such a mechanism can efficiently increase the reward compared to the first scenario, where in the example it is clearly shown that in the case of collusion resistance an increase of 12.46% in the aggregated reward can be achieved.

2.8 Simulation results

In Matlab, we simulate the bandwidth resource allocation in an ISP network, from this network we will study the bandwidth resource allocations over a certain link, utilizing a special pool of 600 Mb bandwidth resources. In this, we compare the resource allocation using both the *Auction Mechanism* (2.2) model and the FAFA queuing model. The models are simulated with 1 to 40 VPN operators competing for these limited bandwidth resources. For each VPN operator, the model generates two values in terms of offered price (representing its utility-gain) and the required QoS class, (Price, QoS), these values are generated through a random function, this function provides values that simulate two options: 1) *Honest* (revealing truthful requirements) and 2) *Exaggerating* (revealing exaggerated requirements), to emulate the operators' options. Traditionally, *FAFA* algorithm allows the first asking VPN operators that satisfy the system requirements to reserve their required bandwidth resources until the resources' pool reaches its maximum limit. The *auction mechanism* proposes a smart selection process with a cost-unit bases to select the cost-efficient operators among the competing VPN operators.

In ARMM, we are expecting to deliver a fair traffic control mechanism that provides better services and higher satisfaction rates for the VPN operators. In addition, we are expecting that such *auction mechanism* would enhance the allocation process performance using the proposed selection algorithm. Moreover, comparing with the ASA FAFA model, we are expecting to enhance the ISPs profit rates by providing them with higher profits resulting from better utilization of their networks' resources. Based on the above assumptions, we construct the following analysis.

Figure 2.2 shows the percentage of satisfied VPN operators to the number of VPN operators participating in the resource allocation game. In this figure, it is shown that VPN operators from 1 to 10 are all satisfied by their provided services, both models have high participation rates. However, as the number of VPN operators increases the required bandwidth resources also increase, where beyond 10 participants, results show that deploying the auction mechanism provides higher satisfaction rates compared with that provided by the FAFA mechanism.

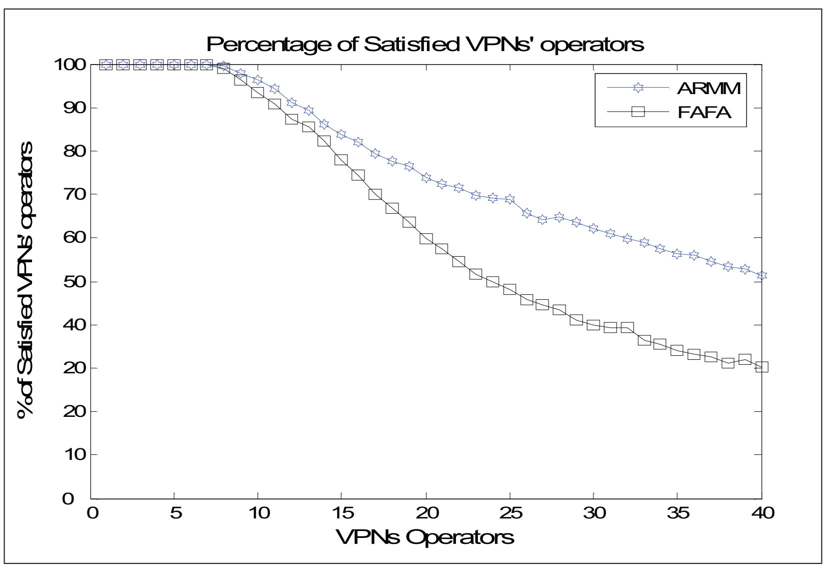


Figure 2.2 Percentage of Satisfied Operators

Figure 2.3 shows the total bandwidth utilization to the number of VPN operators over the considered link. In this figure, it is shown that on average, the auction mechanism provides better usage of the bandwidth resources compared with that provided by the FAFA mechanism, where in the first case winning operators are chosen according to a selective algorithm based on their offered cost-units.

Figure 2.4 shows the total ISP's profit represented by the profit units collected from resources being allocated to VPN operators. In this figure, it is shown that the auction mechanism provides higher profit rates compared to that provided by the FAFA mechanism. Such results provide for longer surviving in the market combined with higher profit rates, where with higher satisfaction rates and better utilization of the available resources, ISPs will be able to provide resources for more VPN operators with competitive prices, which will result in profit rates maximization.

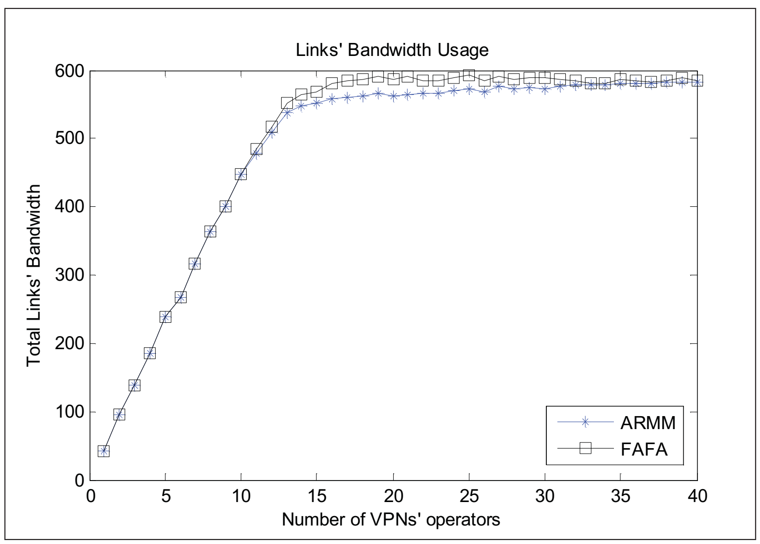


Figure 2.3 Links' Resource Utilization

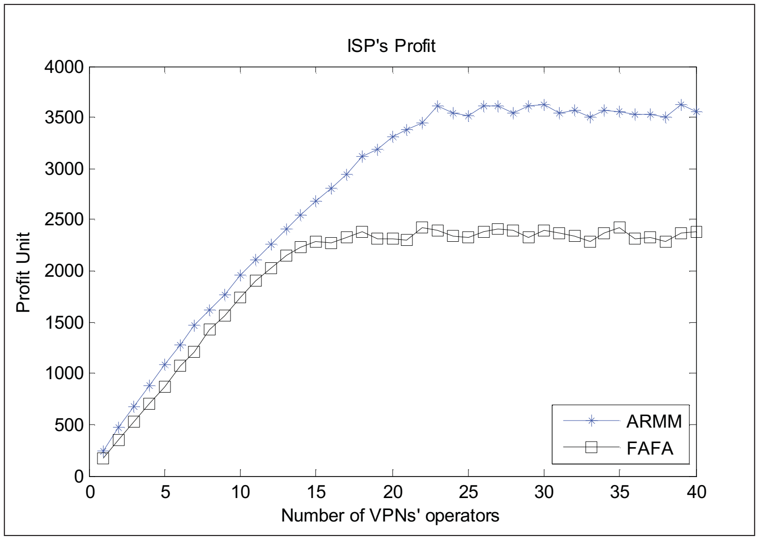


Figure 2.4 ISP's Profit

2.9 Related Work

This section reviews the related work on autonomic resource management and mechanism design applications.

2.9.1 Autonomic Resource Management

The idea of autonomic resource management have been first introduced in Appleby (2001) where the authors developed a prototype of a highly manageable infrastructure for an e-business computing. Their aim was to develop and design a manageable, scalable computing infrastructure that consists of a farm of massively parallel, and densely packaged servers interconnected by high-speed, switched LANs. The concept of dynamic resource allocation was developed to accommodate both planned and unplanned fluctuation of network state under the constraints of the contracted SLAs. In HP-Laboratories (2003), Graupner *et al.* (2003), the vision for the Adaptive Enterprise and the Microsoft Dynamic Systems initiative that are related industry institutions, realizes that autonomic management for the computing components is critical for future Information Technology (IT) industry, Microsoft-Corporation (2004). Also, they both gave emphasize to the importance role of the Virtualiztion concept of network resources. In

Farha and Leon-Garcia (2006), and Sup *et al.* (2005), the authors expanded the autonomic view to include the computer telecommunication services that consider both computing and networking resources. In their work, they proposed the Autonomic Service Architecture (ASA), which is a framework for automated management of both Internet services and their underlying network resources. For this framework, they design an Autonomic Resource Broker (ARB) to serve as the autonomic manager, which is the key enabler of the ASA Farha and Leon-Garcia (2006). The authors in Cheng *et al.* (2006a) proposed a bandwidth sharing scheme for utilizing the available network resources. The bandwidth borrowing scheme provides a way to adapt bandwidth resources that are already allocated for each SLA to be automatically adjusted according to the network state, under defined policies' control for better utilization and QoS guarantees. Finally, in Belzarena *et al.* (2009) the authors addressed the problem of bandwidth allocation via periodic auction. In this work the authors use the first bid auction to allocate the resources where such an auction mechanism cannot eliminate exaggeration, which can affect negatively the ISP revenue. The authors used MDP to determine the allocated optimal prices. Our work differ from this work in the sense that we are using the second price auction (VCG) that can punish exaggerated operators and MDP to avoid collusion.

Although the above studies proposed solutions that can provide an automatic management scheme and dynamic resources allocation, we can note that all the proposed models suffer from the same problem which is customers' exaggeration. Customers' might exaggerate and ask for more unneeded resources whenever this behavior can increase their own revenue. Such a behavior can lead to an inefficient resources allocation scheme and reduce ISP revenue since less resources can be allocated to other customers.

2.9.2 Mechanism Design Applications

Game theory has been proposed in prior research to resolve competitive resource allocation issues for wireless networks in a distributed and scalable manner Kelly (1997a), Semret *et al.* (2005), Badia *et al.* (1997). In Kelly (1997a), a pricing mechanism is adopted for resource allocation to ensure that the sum of users' utilities is maximized. However, the users are assumed to be "price takers" (i.e., they do not anticipate the impact of their actions on the network). In Semret *et al.* (2005), it has been shown that resource allocations such as those proposed in Badia *et al.* (1997) suffer from an "efficiency loss" if the users exploit the fact that their actions affect the network prices. In Semret *et al.* (2005), the auction mechanism was deployed for resource allocation. The optimal auction strategies for the resource-buyers are derived and the equilibrium is shown to exist. In Badia *et al.* (1997), pricing schemes are introduced which could be deployed by a service provider to organize the network. However, the relationship between the assigned resources and gained utility is not thoroughly studied.

In Jeffrey *et al.* (1994), authors proposed an approach that claimed to deliver economic efficiency results. They presented a pricing model that achieves incentive-compatible state, where for each served packet the winner bidder pays a clearing price defined by the highest not-accepted bid (bids of dropped packets). Although this approach might motivate selfish bidders to reveal their truthful information, but on the other hand, it could be classified as not budget-balanced approach, in which, ISPs might provide incentives more than what they actually gain. Authors in Kelly (1997b) proposed an approach that claimed to deliver an efficient pricing model. Their pricing model was built based on a combination of measured and declared traffic characteristics, accordingly an appropriate traffic model is assumed, and corresponding pricing plan is indexed to encourage the bidders to truthfully reveal their information. Although this approach might enhance the resource management process, provide higher profit rates, but it does not address the problem of calculating the optimal prices that bidders have to pay according to the market state. Therefore, we can conclude that this approach does not provide optimal pricing solution.

In Lazar and Semret (1999), a different auction mechanism is proposed to be applied in both additive and arbitrarily divisible resource models. In this work, the proposed model does not consider the QoS as a parameter, where they only assume selling network resources (i.e. bandwidth). This mechanism defines bidders as having an explicit pricing scheme of offered resources, final selling prices are determined based on the Progressive Second Price (PSP) auction mechanism. Briefly, in PSP bidders submit their bids in terms of (required bandwidth and price), then based on the feedback received from the market, bidders reply to any rival bid by manipulating their offers either by increasing the price for the same bandwidth amount or keep the same price but asking for less bandwidth resources, this scenario keeps running until higher offers raised. Consequently, each winner bidder gains the requested bandwidth amount and pays the social opportunity cost (second best price). However, we see that such a mechanism might deliver non-fair allocations, as it could have more than one truthful-equilibrium point.

The authors of Mohamed *et al.* (2011) have used mechanism design to motivate the selfish nodes to cooperate and reveal truthfully their private information. Incentives are given to nodes in the form of reputation where the reputation is calculated based on VCG. Nodes are granted services according to their reputation thus all nodes are now motivated to cooperate and reveal truthfully their information in order to increase their reputation value.

The authors in Dramitinos *et al.* (2007) proposed a mechanism that consists of simultaneous Multi-unit Dutch Auctions (MIDAS) for auctioning bandwidth resources over a wide-bases network, to users that will utilize the resources over the same period of time. Links' price-units are assumed to be asymmetric, reflecting the different demands over the various links. In their work, authors proposed an efficient price reduction policy called the Price Freezing (PF) policy, in which the model deals with the bandwidth pricing problem according to the dynamic bandwidth allocation demands over the considered links or (paths). They assumed that bidders are truth-telling; presenting a payment rule of the VCG type that complements their proposed mechanism and enforces incentive-compatibility and thus truthful bidding. They also proved that their mechanism provides a promising approach to a hard problem, where it requires a low computational complexity and is scalable for large number of units and network users. Never the less, their attained social welfare is close to optimal in general.

In all the above work, mechanism design was used as an incentive mechanism while in our work we have used mechanism design as a threat mechanism. Thus, we are motivating the players to reveal the truthful information under the threat of punishment while in the other work players are motivated by giving them some incentives. Our paper is an extension of Quttoum *et al.* (2010a) where we presented the secure resource allocation without considering customers' collusion. One drawback of the model presented there is that customers could collude with each others to get the resources for less price. Such a behavior can affect ISP profit. This raises the needs for a collusion resistant model that can grant the ISP profit. In this paper, we present such a model based on the MDP and shadow price concepts.

2.10 Conclusion

ASA is an interesting model that was proposed in order to have an autonomic resource utilization scheme. However, this model has different limitations such as the non-optimal increase of the ISPs' revenue and the lack of an incentive scheme that can cope with customers' exaggeration, especially in the case of multimedia networks. To optimally increase the ISPs' revenue and to avoid customers' exaggeration, we proposed to utilize resources through the auction scheme and to eliminate exaggeration behavior by the use of mechanism design. Here, we modeled a threat mechanism based on VCG that penalizes customers according to the amount of inconvenience created. Our example showed how exaggerated customers are charged high transfer rates that affect negatively their revenue rates. We assumed that customers are rational, therefore they will avoid such type of behavior but at the same time they are always motivated to cheat if such a behavior can increase their benefits. Collusion is one of the tactics that VPN operators can use to have the resources with less prices. This can maximizes customers' payoff and reduces ISP revenue. To overcome such a problem, an optimal price threshold was modeled based on the MDP and shadow price concept. The shadow price concept was used to calculate the optimal cost-unit at every new state, where bids that are greater than the threshold are considered during the auction game. Such a threshold is able to eliminate the effect of collusion on ISP profit and the VCG model is able to eliminate the VPN operators' exaggeration. Our results showed that our model is able to increase ISPs' revenue, satisfy more customers and utilize efficiently the network resources.

CHAPTER 3

AN OPTIMAL DYNAMIC RESOURCES PARTITIONING AUCTION MODEL FOR VIRTUAL PRIVATE NETWORKS

Ahmad Nahar Quttoum[1], Abdallah Jarray[1], Hadi Otrok[2], and Zbigniew Dziong[1]

[1]Electrical Engineering Dep., École de Technologie Supérieure, Université du Québec

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2]Computer Engineering Dep., Khalifa University of Science, Technology & Research, UAE

An Article submitted to the Springer Journal of Telecommunication Systems in May 2011

3.1 Abstract

In this paper, we consider the problem of optimizing the Internet Service Provider (ISP) profit by providing a periodic Dynamic Partitioning (DP) model for utilizing network resources in the context of Virtual Private Networks (VPN). In literature, Complete Sharing (CS), Complete Partitioning (CP), and Bandwidth Borrowing (BR) techniques have been proposed for resource allocation where the following limitations can be noticed: VPN operators can exaggerate about their required resources, resources might be underutilized, and optimal bandwidth utilization is not guaranteed. To overcome the above limitations, we propose to dynamically partition the resources over different QoS classes through periodic auctions that can reduce the reasoning of exaggeration and maximize the ISP profit. Thus, we formulate our problem based on the Integer Linear Programming (ILP) that allows us to maximize the ISP profit and provides the optimal: (1) set of profitable VPN connections, (2) bandwidth division of each network link among QoS classes, and (3) routing scheme for the accepted demand. Furthermore, the proposed ILP model allows us to study the sensitivity of the ISP profit to a targeted revenue objective.

3.2 Introduction

We are witnessing an unprecedented demand for Internet Service Providers (ISPs) to support a wide variety of Virtual Private Networks (VPNs). These VPNs use the ISPs' public infrastructure to establish secure and reliable services according to contracted Service Level Agreements (SLA), Duffield *et al.* (1999a). Resource management is one of the main challenges facing ISPs, where each VPN may have different requirements and required Quality of Service (QoS) guarantees. The growing demands for such VPN services with certain QoS satisfaction conditions necessitate the ISPs to efficiently utilize their bandwidth resources. Models to provide an efficient and effective resource allocation schemes are extremely important. Relying on the current management models that attempt direct interactions with the ISPs is not satisfying anymore, as they have many limitations that can be summarized as follows: 1) These models increase the management operation expenses. 2) They provide slow response times. Such limitations can lead to high rates of users' dissatisfaction. Consequently, the need is emerging to find an alternative resource management models that overcome the current limitations. Complete Sharing (CS), Lai *et al.* (1998); Mark *et al.* (2000), and Complete Partitioning (CP), Lai *et al.* (1998); Haung and Ho (2002); Cheng *et al.* (2005), models are proposed in the literature to cope with the above management problem by creating a framework for automated management. In CS, network resources are shared over all classes without any division. While, in CP, resources are statically divided among QoS classes where each class uses its own allocated resources. Such approaches are able to automate the resources allocation, while at the same time they create several problems that can be summarized as follows:

- in CS approach, resources are allocated in a First Ask First Allocate (FAFA) scheme, where one QoS class can overwhelm all other classes in a way that reduces the ISP profit;
- in CP approach, resources can be underutilized which reduces ISPs profits;
- due to the long-term SLA agreement, VPN operators might exaggerate about their required resources to guarantee their QoS and to cope with any unpredicted variations in

the network state. Such a behavior can affect the ISP profit, and maximize the VPN demands blocking rates. Indeed, when demand for the resources exceeds the capacity, avoiding exaggeration allows to admit more VPN connections (using the same amount of bandwidth resources), which can reduce the blocking rates while collecting the same profits.

To overcome the problem of underutilization in CP, the Bandwidth Borrowing (BR) technique has been proposed, Sup *et al.* (2005), Cheng *et al.* (2006a), to enable the ISPs to provide better resource utilization that guarantees the QoS for all VPNs. According to the dynamicity of the traffic, the BR technique allows the ISP to borrow the extra resources of the underloaded links and reallocate them to the overloaded ones. While yet, as long as the CP attempts a static partitioning scheme, BR cannot provide an optimal bandwidth utilization solution, where we may have many underloaded links with no overloaded ones, in such a case how the BR would work?

To overcome the above limitations, we propose deploying a *periodic-dynamic resource partitioning model* (DP) for allocating the network bandwidth resources to VPN operators. In DP, the dynamic class division process will take place in a periodic auction manner, where VPN operators will be competing to win the resource allocations through their submitted prices and connections' demands. The network planning time is divided into a set of periods in order to reduce the reasoning of exaggeration and to eliminate the needs for any borrowing technique. The length of the re-optimization period depends on the provided service duration. Demands with different service durations can be divided into groups, each with certain period length. At each new period, the profitable VPN connections are selected through an auction mechanism in order to maximize the ISP profit. Therefore, we develop and implement an *Integer Linear Program* (ILP) solved using the ILOG CPLEX concert technology environment. The Proposed ILP maximizes the ISP profit and comes-up with:

- the optimal set of the profitable VPN connections;
- the optimal bandwidth division of each network link among different QoS classes;

- the optimal routing scheme of the accepted VPN connections, with respect to their QoS specifications.

Furthermore, the proposed ILP allows us to study the sensitivity of the ISP profit to a targeted revenue objective represented by the Profit Percentage Parameter (PPP). Through PPP, we are able to find the optimal tradeoff setting that maximizes the ISP profit while not breaking the demands blocking constraint. Consequently, our contribution is a model that is able to:

- efficiently utilize the network bandwidth resources, where through the dynamic partitioning we are able to consider the changes of traffic demands;
- increase the VPN operators' satisfaction rates, since resources are utilized efficiently and the allocation prices are market competing;
- minimize the SLAs violations as exaggeration motivations are reduced.

3.2.1 Paper Organization

The rest of the paper is organized as follows: Section 3.3 presents the problem statement. Section 3.4 illustrates our dynamic resource partitioning model and the proposed selection algorithm. Section 3.5 lists the proposed performance metrics, followed by the computational results in Section 3.6. In Section 3.7, we present the related work. Finally, Section 3.8 concludes the paper.

3.3 Problem Statement

The *Complete Sharing* (CS) approach proposes an autonomic management framework, where the whole network resources are shared between all VPN operators, so each of them can use any bandwidth available at any given moment. With no link bandwidth division, in the CS approach, all QoS classes can share the network resources without discrimination. Through this, the model is supposed to ensure the delivery of services according to predefined SLAs. These SLAs are constructed after negotiations between the ISP and the VPN operators. From the ISP prospective, giving the VPN operators such a privilege can provide better utilization of the network resources, assuming that the VPN operators know well their changing requirements, and accordingly they can acquire the required resources based on their actual needs. Naturally, as long as the VPN operators behave according to their declarations, deploying such an approach can increase bandwidth utilization due to the statistical multiplexing, which results in better satisfaction rates and higher profits at the same time. However, in reality, VPNs can exceed their declarations as long as there is no link bandwidth division, and one class may overwhelm all other classes which creates several problems like SLAs violations, high blocking ratios, and low profit rates.

In the *Complete Partitioning* (CP) approach, the scenario is somehow different. Here the resources are partitioned among the provided service classes in a *static* way, based on a one-time bandwidth partition for all network links. With such a static scheme, each class is allowed to exclusively use its portion of the provided resources. This can solve the SLAs violation problem, but it may lead to low bandwidth utilization as resources might be underutilized.

Exaggeration is considered as a common drawback of both CS and CP, where VPN operators might exaggerate their requirements in order to cope with any sudden or unpredictable changes in the network state and link conditions, so they tend to keep a spare amount of resources that enables them to overcome and cope with such situations. This is due to the long-term SLA that does not allow users to change their resource requirements according to their needs. Such a problem can be solved either through a short term resource sharing model that reduces the reasoning of exaggeration or through a penalization model that penalizes any exaggeration behavior,

Quttoum *et al.* (2010a), and Quttoum *et al.* (2010b).

To overcome the problem of resource underutilization, the *Bandwidth Borrowing* (BR) technique has been proposed, Cheng *et al.* (2006a). The BR attempts the idea of borrowing the resources from the underloaded links and allocating them to the overloaded ones. Accordingly, the BR scheme can provide better utilization of the network bandwidth resources, and better

SLAs guarantees. However, resources are utilized in a non-optimal way that can reduce ISP profit. Also, such behavior can entail a high management load to cope up with the changes of traffic load and utilize the unused resources in realtime state.

3.4 Dynamic Resource Partitioning Approach

In this section, we present our *Dynamic Partitioning* (DP) approach that improves the CS and CP techniques, and overcomes their limitations by virtually partitioning the network bandwidth resources in an efficient way based on *periodic auctions*, where resources are allocated to the best bidders providing profit maximization. In this context, we propose a resource management approach based on short-term SLAs that takes into account the desired ISP objectives over the time. To achieve this, VPN operators are asked to reveal their connection demands in terms of: source-destination nodes, required QoS classes, and their offered prices (bids). Accordingly, Algorithm 3.1 is deployed to provide an optimal allocation mechanism in the DP model.

3.4.1 DP Approach

3.4.1.1 The DP Algorithm

In our DP approach, resources are allocated based on *dynamic auctions* that are held in a periodic manner, with the aim of selecting the best set of VPN operators from a competing environment. VPN operators' requests are represented in terms of *connection demands*. The term connection here is defined as a secure network tunnel that is layered on top of a public network, such tunnel has some QoS parameters (i.e. bandwidth capacity, maximum number of hops), and it is used to send data from a source node s_k to another destination node d_k. The process of selecting the winning set of VPN connections for resource allocations deploy an optimal selection procedure presented in Algorithm 3.1. In which, at each auction period t, the ISP broker firstly collects the new connection demands received in the period of $(t-1, t]$ and form a new demand matrix, then from each demand matrix it chooses the most profitable set of connections to accept. To do so, using the Linear Programming Theory we developed an Integer Linear Program (ILP) to formulate the bandwidth allocation problem. By solving this ILP, we are showing how it can choose the optimal set of bidding VPN operators, and provide the optimal bandwidth utilization rates. The ILP checks the offered bids of the competing VPN operators if it is larger than or equal certain selling-price thresholds, where such thresholds depend on the ISP broker profit objectives, the corresponding QoS classes, and demand blocking constraints. Demand blocking is mainly related to the number of offered bids, and the bandwidth availability over the assigned candidate paths.

Algorithm 3.1 DP Selection Algorithm

Selecting the Winning VPN operators in DP

1: **Input:** At each auction round t, the ISP broker **do**;
2: Collect the VPN operators connection demands K (p_k, QoS_k, s_k, d_k) received through time $(t-1, t]$;
3: Formulate the problem as an ILP;
4: **Solve** the ILP, and find the optimal set of VPN connections;
5: **Output:** the VPN connections that won the resource allocations and the profit collected from these connections;

To overcome the problem of exaggeration, we propose performing a dynamic partitions of resources over different QoS classes through short periods of time comparing with that of the static scenarios. This can reduce the motivations of exaggeration actions, where bidding VPN operators will not be motivated anymore to ask for more resources, as they do not have to care about their future connections or the sudden changes in the network state, at least at this stage, since current allocations are only valid for a short period of time, comparatively.

3.4.1.2 Mathematical Modeling

We modelize the DP appraoch as an ILP model. Through this model, we study the case where we assume having a number of VPN operators N competing for bandwidth allocations over a network that is managed through an ISP broker. Bandwidth allocation requests are modeled through connection demands K, where each VPN operator i, $i \in N$, has a set of connection demands k belonging to a set of service classes J. Logically, VPN operators are assumed to be rational, and thus their aim is to have their connection demands admitted with the lowest possible prices, and at the same time acquire satisfactory QoS levels. On the other side, the ISP broker aims to utilize its network bandwidth resources better, and maximize its profit by accepting the maximum number of VPN operators while providing them with satisfactory QoS levels by competing prices. Accordingly; the ISP broker profit P_{B} function can be expressed as:

$$P_{\mathrm{B}} = \max \left(\sum_{k=1}^{|K|} p_k - \sum_{k \in K} c_k \right) \tag{3.1}$$

where the P_{B} equals the sum of bids p_k collected from connection demands being admitted for allocation, reduced by the total cost of the bandwidth resources required to satisfy each connection demand c_k. Accordingly, to maximize the P_{B}, the ISP broker has to choose the best set of VPN operators' connections that maximizes the first term of the function, and also, deploy an efficient bandwidth utilization scheme in order to minimize the second term.

Choosing the best set of VPN operators' connections is typically done in accordance to their offered bids, where higher bids increases the chance for their corresponding demands to be accepted. To create a form of competition between the bidding VPN operators, we assume that the amount of bandwidth resources required to satisfy the received connection demands exceeds the available network capacity. Connection demands are also classified into set of different QoS classes summarized in Table 3.2, where for each auction period t, the model receives a different set of connection demands' patterns reflecting the diversity of the required services at the different auction periods. Periods can vary between day and night times, weekdays, and weekends.

Table 3.2 QoS Classes' Map

QoS Class	Connections Quality of Service	Bandwidth per Connection	Max. of Hops
5	Golden Load (<100ms Latency)	5Mbps	2 Hops
4	Excellent Load (Business Critical)	4Mbps	3 Hops
3	Controlled Load (Streaming Video)	3Mbps	4 Hops
2	Standard (IP Packet Delivery)	2Mbps	5 Hops
1	Best Effort	1Mbps	6 Hops

Model Parameters:

The ISP network is basically defined as a set of nodes V, connected by a set of bidirectional links L, where each physical link offers certain bandwidth capacity b_l. The requests of the VPN operators N form the shape of the demands matrix K, where K represents all the connection demands submitted by the bidding VPN operators, each demand consists of a VPN operator ID i, connection source node s_k, destination node d_k, and a class of service j, $j \in J$. Henceforth, a connection demand k of class j, consumes a bandwidth amount of b_j. Dividing the links' bandwidth capacities among the set of connection demands K and their corresponding service classes J highly depends on the period's traffic pattern. β_j^l denotes the percentage of bandwidth capacity allocated to a connection of class j over the network link l, where $\beta_j^l \leq b_l$.

Model Variables:

To facilitate the process of measuring the ISP broker utility presented in Equation 3.1, we refer to the admitted/rejected connections using the variable z_k, where:

$$z_k = \begin{cases} 1 & \text{if connection demand } k \text{ is admitted} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

and we also refer to the selected path that holds the considered connection using the variable x_k^π, where:

$$x_k^\pi = \begin{cases} 1 & \text{if connection demand } k \text{ uses path } \pi \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

Objective Function:

According to the parameters and variables defined above, the objective function of Equation 3.1 can be reformulated as:

$$P_{\mathrm{B}} = \sum_{k \in K} \left(p_k z_k - \sum_{\pi \in \pi_k} x_k^\pi \sum_{l \in \pi} c_k^l \right) \tag{3.4}$$

Model Constraints:

For this objective function presented in Equation 3.4, we have the following constraints:

- Link Capacity: For each link l, the available bandwidth for the admitted connection demands k of class j cannot exceed the total link capacity reserved for this class.

$$\sum_{k \in K_j} \sum_{\pi \in \pi_k} \eta_l^\pi x_k^\pi b_j \leq \beta_j^l b_l \qquad ;\ l \in L,\ j \in J \tag{3.5}$$

where the variable η_l^π refers to:

$$\eta_l^\pi = \begin{cases} 1 & \text{if path } \pi \text{ uses link } l \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

and K_j refers to the set connections belonging to QoS class j .

- Link bandwidth Division among QoS classes: The summation of the defined class divisions over the network links must not exceed the total link capacity.

$$\sum_{j \in J} \beta_j^l = 1 \qquad ; \beta_j^l \in [0,1],\ l \in L \tag{3.7}$$

- Minimum Selling-Price Threshold: To be considered as a competing connection demand for the allocation process, the minimum offered bid p_k for a connection demand k of QoS class j must at least be higher than or equal to a certain threshold. The calculation of this threshold is derived in subsection 3.4.1.4.

$$p_k \geq \sum_{\pi \in \pi_k} x_k^\pi \sum_{l \in \pi} p_{th,j}^l \qquad ; k \in K_j,\ j \in J \tag{3.8}$$

- Routing Path Assignment: Only one routing path can be assigned to carry each connection.

$$\sum_{\pi \in \pi_k} x_k^\pi \leq 1 \qquad ; x_k^\pi \in \{0,1\},\ k \in K \tag{3.9}$$

- Linking decision variables: Finding a path for the connection demand does not guarantee final acceptance.

$$z_k \leq \sum_{\pi \in \pi_k} x_k^\pi \qquad ; z_k \in \{0,1\},\ k \in K \tag{3.10}$$

- Routing Path Length: The length of the assigned path l_π for a connection demand k of QoS class j cannot exceed H_j hops, one hop represents one physical link.

$$l_\pi \leq H_j \qquad ; \pi \in \pi_k,\ j \in J \tag{3.11}$$

ILP Complexity:

However, it is worth to find the computational complexity of the model. This can be measured through the number of model variables and constraints, as follows:

- Number of variables:

$$|K| + |K| \times |\pi_\mathrm{K}| + |L| \times |J|$$

which can be simplified as:

$$|K| \times |\pi_\mathrm{K}| + |L| \times |J|$$

- Number of constraints:

$$|L| \times |J| + 3|K| + |K| \times |\pi_\mathrm{K}|$$

which can be simplified as:

$$|L| \times |J| + |K| \times |\pi_\mathrm{K}|$$

From this, we can conclude that the number of variables equals the number of constraints, which is in the order of $O(n^2)$. Moreover, to deal with the complexity issues, if exist, in such a problem we can use the technology of *Cloud Computing* that is considered as an efficient solution to deal with high combinatorial problems.

3.4.1.3 Link Bandwidth Division Scheme

To assign the VPN operators' demands to the most profitable paths, the ILP do the following:

- in accordance to the offered bid rates and the network links capacities, it generates a class division map that shows the percentages of bandwidth capacities reserved for each QoS class over the network links;
- based on this map, the model assigns different candidate paths, Epstein (1994), to carry each source-destination connection demand;
- in one-shot scheme, the model chooses the most profitable combinations of the VPN operators' demands to be carried over the network links, and consequently assign them to their routing paths.

To create a form of competition, the assigned candidate paths might be shared between various connection demands having the same source-destination nodes, where the sum of bandwidth amounts over these candidate paths is less than or equal to that required to satisfy the whole connections with the same source-destination couples.

The link bandwidth division among QoS classes scheme mainly depends on the demands' diversity, where such diversity reflects the different VPN operators requirements from time to time. Differences may appear between the day times (i.e. mornings, afternoons, evenings, and nights), weekdays (i.e. workings days, and weekends), or even it might be affected by the different time-zones. Consequently, for each period of time we assume having the following demand matrix partition:

- $\gamma_1\%$ connection demands of QoS 1;
- $\gamma_2\%$ connection demands of QoS 2;
- $\gamma_3\%$ connection demands of QoS 3;
- $\gamma_4\%$ connection demands of QoS 4;
- $\gamma_5\%$ connection demands of QoS 5.

3.4.1.4 Profit Percentage Parameter

For each VPN connection k belonging to QoS class j, we calculate the set of candidate paths π_k. Then, if at least one path π, $\pi \in \pi_k$, uses the link l, then we add the VPN connection k to the set of candidate VPN connections that use the class of service j over link l. We denote by K_l the set of candidate VPN connections to use the link l:

$$K_l = \sum_{j \in J} K_l^j \tag{3.12}$$

where K_l^j is the set of VPN connections belonging to QoS class j, candidates to use the link l. Accordingly, for each link l, we can define ratio β_j^l of bandwidth allocated to the use of class j VPN connections by:

$$\beta_j^l = \frac{b_j |K_l^j|}{b_l} \tag{3.13}$$

Having the allocation percentage β_j^l and the original cost-unit c_l of link l, we can define *threshold* $p_{thr,j}^l$ as:

$$p_{th,j}^l = b_j c_l + b_j c_l (a \beta_j^l) \tag{3.14}$$

where c_l is the bandwidth original cost-unit over the link l, and the parameter a refers to the *profit percentage parameter* (PPP). This PPP reflects a *tradeoff relationship* between the ISP broker profit objective and the demand blocking ratios. Consequently, as long as the demand blocking ratio B_K is less than the blocking constraint B_K^c, $B_K < B_K^c$, optimal ISP broker profit can be achieved by assigning higher a values. Additionally we assumed in equation (3.14) that the profit is proportional to percentage β_j^l. That means that the value of the threshold is larger for the QoS classes that use more bandwidth. The motivation behind this scheme is as follows:

- it reduces the likelihood of exaggeration. Indeed, with such a scheme, the VPN users will be motivated to use the lowest amount of bandwidth resources that satisfy their needs, especially is the QoS class that use a lot of resources;
- it limits situations where one QoS class overwhelms the others. This provides a kind of fairness between the classes.

Obviously, in general, one can substitute β_j^l in (3.14) by another function of system parameters (being one in the simplest case) to realize particular objectives of given ISP.

Results in Subsection 3.6.2 show the tradeoff relationship between both blocking ratios and ISP broker profits for different values of the PPP a. Accordingly, Equation 3.4 can be reformulated as:

$$P_{\mathrm{B}} = \sum_{j \in J} \sum_{k \in K_j} \left(p_k z_k - \sum_{\pi \in \pi_k} x_k^{\pi} \sum_{l \in \pi} p_{th,j}^{l} \right) \tag{3.15}$$

where bids accepted for the competition must be larger than or equal the candidate path *minimum price threshold*.

3.4.2 Benchmark Models

In DP, the dynamic class division process will take place in a periodic auction manner, where VPN operators will be competing to win the resource allocations through their submitted prices and connections' demands. The network planning time is divided into a set of periods in order to reduce the reasoning of exaggeration and to eliminate the needs for any borrowing technique. Thus, there is no need for such a BR technique since there is no underloaded links as result of exaggeration. To evaluate our model, we selected to adapt the CS Cheng *et al.* (2006a) and CP Lai *et al.* (1998) models to run periodically.

3.4.2.1 CS Model

In this technique, in addition to the poor QoS guarantees, the VPN connections selection is non-efficient since the process of allocating the network bandwidth resources is done according to a First Ask First Allocate (FAFA) scenario. Accordingly, new VPN operators can guarantee the resources for their connection demands as long as: 1) Their offered bids p are larger than or equal certain selling cost thresholds. 2) At their arrival time, there exist enough bandwidth resources to hold their connection demands over the considered paths. Algorithm 3.3 illustrates the allocation mechanism in the CS model.

Algorithm 3.3 CS Selection Algorithm

Selecting the Winning VPN connections in CS

1: **Input:** VPN operator submit its connection demand
$k \rightarrow$ ISP broker where, for each k
we have (p_k, QoS_k, s_k, d_k) , QoS_k=(b_k, H_k);
2: Sort the connection demands k according to their **arrival** sequence;
3: **for** each demand k, **do**;
3.1: Find the set of candidate paths π_k to hold its connection where each $\pi \in \pi_k$, uses less than H_k hops (i.e., links);
3.2: For each π, $\pi \in \pi_k$ check if:
$\forall l \in \pi$, $b_k \leq b_l(residuel)$;
$p_k \geq \sum_{l \in \pi} p^l_{th,QoS_k}$;
3.3: If $\pi_k \neq \varnothing, then$:
3.4: Choose one random path π to hold the connection k ;
3.5: else, reject;
4: **Output:** the set of VPN connections that won the resource allocations and the ISP profit collected from these
these connections;

In the first step, for each connection demand k, the corresponding VPN operator is asked to submit its offered bid p_k, required QoS_k commitments, i.e., bandwidth b_k and maximum number of hops H_k, and source-destination nodes (s_k, d_k) to an ISP broker in order to consider its requests. In step two, the ISP broker sorts the connection demands according to their arrival sequence. In step three, for each connection demand k, the ISP broker finds the set of candidate paths π_k to hold the connection (shortest paths), checks their bandwidth availability and the corresponding offered bid price. Next, the ISP broker verifies that there exists at least one candidate path to hold the connection. Once the bandwidth availability and the bid price conditions are satisfied, the broker choose randomly, Cheng *et al.* (2006a), one path from the set π_k and assign it to hold the connection of the corresponding VPN operator. Last, in step four, the ISP broker calculates the profit collected from the admitted VPN connections according to their offered bid prices.

Consequently, in CS, the ISPs broker profit function P_{B} can be represented by:

$$P_{\mathrm{B}} = \sum_{k=1}^{k_{\max}} (p_k - \sum_{\pi} \sum_{l \in \pi} c_{\pi}^{l}) \tag{3.16}$$

where it equals the sum of bids collected from the accepted VPN operators' connection demands k, $(k = 1, 2, , ..., k_{max})$, given that k_{max} is the index of the last admitted demand (sorted in a FAFA order) that can fit within the available network bandwidth resources, reduced by the cost of resources c_{π}^{l} used on the assigned routing paths. Also each of the accepted bids has to fulfil the following requirements: 1) The chosen path to hold the VPN operator connection k must at least meet the QoS_k commitments, in terms of the allocated bandwidth b_k and the maximum number of hops H_k. 2) Total bandwidth amounts allocated to the accepted connections cannot exceed the sum of physical bandwidth capacities of the network links. 3) Accepted bids must at least be larger than or equal certain selling-price threshold, associated with the required QoS_k class at the chosen path π to hold the connection.

3.4.2.2 CP Model

The process of selecting the winning set of VPN connections for resource allocations follows the selection scheme presented in Algorithm 3.3 except we add the following modifications. As aforementioned previously, the link bandwidth division among QoS classes is done at the beginning of the network planning (period 0) based on the shortest path algorithm, and the traffic pattern statistics. In other words, we fix the variables β_j^l to the resulting value calculated at period 0. The resulting link partitioning among QoS classes is used for the rest of planning periods.

3.5 Performance Metrics

As mentioned in the previous section we propose to use CS and CP as benchmark models to evaluate the performance of the DP approach. For each model we are measuring the following metrics:

3.5.1 ISP Broker's Profit

The ISP broker's profit, Equation 3.4, is measured based on the bids collected from the admitted connection demands, reduced by the cost-unit of the carrier network links. The bids and the cost-units are expressed in terms of $\$X$, which represents the price of 1Mb of bandwidth. This metric represents the profits collected from the whole bandwidth resources allocated over the network links.

3.5.2 ISP Broker's Profit-Unit

The profit-unit value represents the gains collected per $1Mb$ of bandwidth, which is measured as the ratio between the ISP borker profit, calculated in Equation 3.4, and the amount of bandwidth over the whole network links, as:

$$P_{\mathrm{B}}^{u} = \frac{P_{\mathrm{B}}}{\sum\limits_{l \in L} b_l} \tag{3.17}$$

3.5.3 Demands' Blocking Ratio

Demands' blocking ratio represents the VPN operators' satisfaction rates, where it is measured as the ratio between the number of admitted connection demands to the number of the whole demands participated in the allocation auction, as

$$B_K = \frac{\sum\limits_{k \in K} z_k}{|K|} \tag{3.18}$$

3.5.4 Bandwidth Utilization

Bandwidth Utilization is measured as the ratio between the used and the total bandwidth amounts, as:

$$U_b = \frac{\sum\limits_{j \in J} \sum\limits_{k \in K_j} b_j z_k}{\sum\limits_{l \in L} b_l} \tag{3.19}$$

3.5.5 Routing Scheme Efficiency

To evaluate the routing scheme efficiency of the above mentioned scenarios, we proposed measuring the following parameters:

3.5.5.1 End-to-End Delay

The average end-to-end delay per admitted connection demand is measured through the average number of *hops* used to form a routing path, which is given by the ratio between the number of hops counted in composing the whole routing scheme to the number of admitted demands. Only routing paths of the admitted connection demands are counted. This can be formulated as:

$$H = \frac{\sum\limits_{k \in K} \sum\limits_{\pi \in \pi_k} x_k^{\pi} l_{\pi}}{\sum\limits_{k \in K} z_k} \tag{3.20}$$

3.5.5.2 Number of Used Network Links

In this we measured the number of *links* used to form the whole routing scheme in the assigned period of time.

3.6 Computational Results

To assess the efficiency of this work, in this section, we illustrate an example of a network that consists of 10 nodes, connected by 40 bidirectional links, for which we assume that each link holds a $25Mb$ of bandwidth resources. For this network we assume receiving 100 VPN operators' connection demands per period of time, and the network planning time is divided into 12 periods per week in a way that shows the main periods of a $24 - hour$ day, Thompson *et al.* (1997), representing a new allocation auction every 6 hours. However, the length of the re-optimization period depends on the provided service duration.

Demands with different sevice durations can be divided into groups, each with certain period length. In this work, we assume that the ISP broker is selling certain *service packages* that stands for 6 hours of time, where there is no dependency between any two consecutive periods and each period starts by a new auction. Such service packages can vary between a VoIP call, chat session, an IPTV connection, an on-line game, video streaming or even a movie show. Table 3.4 below shows the periods' division map. Through this network, we will study the behavior of the above mentioned three scenarios (DP, CP, and the CS FAFA).

Table 3.4 Periods' Division Map

Period Index	Corresponding Time Period
1	Monday to Friday: Morning
2	Monday to Friday: Afternoon
3	Monday to Friday: Evening
4	Monday to Friday: Night
5	Saturday: Morning
6	Saturday: Afternoon
7	Saturday: Evening
8	Saturday: Night
9	Sunday: Morning
10	Sunday: Afternoon
11	Sunday: Evening
12	Sunday: Night

3.6.1 Numerical Results

In the studied scenarios, different bandwidth division schemes are followed. As mentioned before, the *CS* model follows a FAFA allocation scheme, where there is no link bandwidth division among QoS classes at all. In *CP*, the model uses a static bandwidth division scheme. At period 0, the ISP broker attempts a heuristic link partitioning based on the shortest path algorithm and the traffic pattern statistics. The resulting divion of link bandwidth among QoS classes is used for the rest of planning periods. On the contrary, in DP approach a dynamic link bandwidth division process is performed at each period, and it is handled by the ILP in the *Optimal* manner. The provided QoS classes, their reserved bandwidth amounts, and their corresponding profits at three different periods are presented in Tables 3.5, 3.6, and 3.7. To improve the readability of the results, the resulting profit-units of both CP and DP are also given. Through the presented tables, we are showing the effect of the different *link bandwidth division schemes* used by the three studied scenarios. At each case, we compare the resulting total profits and the corresponding profit-units.

For better comparison and to show how our model can deal with different traffic patterns, in each table, we present the results of the afternoon period at the Working weekdays, Saturdays,

Table 3.5 Bandwidth Class Division, Bandwidth Amounts Allocated and corresponding Total and Per-Unit Profits in three scenarios
(Scenario A: Complete Sharing, B: Complete Partitioning, C: Dynamic Partitioning)
(Profit Percentage Parameter $a = 1.3$, Period = 2)

QoS Class	*CS: Complete Sharing*		*CP: Complete Partitioning*		*DP: Dynamic Partitioning*		*Profit unit in CP*	*Profit unit in DP*
	Bandwidth amount Allocated	*Total Profit in ($X)*	*Bandwidth amount Allocated*	*Total Profit in ($X)*	*Bandwidth amount Allocated*	*Total Profit in ($X)*		
1	*No Division*	157	252	248	251	248	**0.98**	**0.98**
2	*No Division*	260	231	351	228	481	**1.51**	**2.10**
3	*No Division*	391	321	661	262	850	**2.05**	**3.24**
4	*No Division*	98	48	119	25	163	**2.47**	**6.52**
5	*No Division*	267	117	213	226	570	**1.82**	**2.52**

and Sundays. Table 3.5 shows the results of period 2 (Monday to Friday Afternoons). By aggregating the resulting total profits of the different scenarios, and assuming $\$X = \$1$ we collect: \$1173 from the CS FAFA, \$1592 from the CP, and \$2312 from the DP. Results reflect the impact of the deployed *link bandwidth division scheme* on the final profits, where the CS FAFA provides the worst rates, followed by the CP providing an increase of 35%. DP provides the best results, showing an increase of 97% and 45% to that provided by the CS FAFA and the CP, respectively.

Table 3.6 Bandwidth Class Division, Bandwidth Amounts Allocated and corresponding Total and Per-Unit Profits in three scenarios
(Scenario A: Complete Sharing, B: Complete Partitioning, C: Dynamic Partitioning)
(Profit Percentage Parameter $a = 1.3$, Period = 6)

QoS Class	*CS: Complete Sharing*		*CP: Complete Partitioning*		*DP: Dynamic Partitioning*		*Profit unit in CP*	*Profit unit in DP*
	Bandwidth amount Allocated	*Total Profit in ($X)*	*Bandwidth amount Allocated*	*Total Profit in ($X)*	*Bandwidth amount Allocated*	*Total Profit in ($X)*		
1	*No Division*	86	105	125	110	134	**1.19**	**1.21**
2	*No Division*	80	127	138	130	211	**1.08**	**1.62**
3	*No Division*	388	456	935	320	1088	**2.05**	**3.40**
4	*No Division*	224	63	112	98	204	**1.77**	**2.08**
5	*No Division*	752	221	558	336	1244	**2.52**	**3.70**

The same conclusions can be drawn from Tables 3.6 and 3.7 where at the period 6 (Saturday Afternoon) results in Table 3.6 show a \$1530 collected from the CS FAFA, \$1868 from the CP, and \$2881 from the DP by an increase of 88% and 54% to both CS FAFA and the CP, respectively. However, in some cases we see that the CS FAFA may provide higher instant profits like that for QoS classes 4 and 5 in Table 3.6. This results from the random allocation of the resources which might be profitable in some cases. Though, what concerns the ISP is the final profit rates, and experiments show that the resulting total profits of the CS FAFA are always much lower than that provided by the DP schemes. Records of Table 3.7 are also showing that the DP scheme delivers the highest profit rates, providing an increase of 95% and 30% to both CS FAFA and the CP schemes, respectively.

Table 3.7 Bandwidth Class Division, Bandwidth Amounts Allocated and corresponding Total and Per-Unit Profits in three scenarios (Scenario A: Complete Sharing, B: Complete Partitioning, C: Dynamic Partitioning)(Profit Percentage Parameter $a = 1.3$, Period = 10)

QoS Class	*CS: Complete Sharing*		*CP: Complete Partitioning*		*DP: Dynamic Partitioning*		*Profit unit in CP*	*Profit unit in DP*
	Bandwidth amount Allocated	*Total Profit in (\$X)*	*Bandwidth amount Allocated*	*Total Profit in (\$X)*	*Bandwidth amount Allocated*	*Total Profit in (\$X)*		
1	*No Division*	86	105	125	117	134	**1.19**	**1.14**
2	*No Division*	89	110	133	115	184	**1.20**	**1.60**
3	*No Division*	383	491	1002	420	1209	**2.04**	**2.87**
4	*No Division*	692	219	680	280	975	**3.10**	**3.48**
5	*No Division*	122	52	121	50	180	**2.32**	**3.60**

Comparing the three different scenarios shows that the DP scheme always provide the best results in terms of *efficient bandwidth division* and *final profit rates*, followed by the CP and lastly the CS FAFA scheme. It is worth to mention that the DP scheme also provides the lowest *demand blocking ratio*, while it is much bigger for the CS FAFA scheme.

3.6.2 Performance Results

Through this subsection, using the CPLEX optimizer, we studied the behaviour of our proposed DP model compared to both CS and CP in terms of: 1) Bandwidth efficiency. 2) ISP broker total profits. 3) Demand blocking ratios. 4) ISP broker profit-unit, and 5) Routing efficiency in terms of average end-to-end delay, and number of used links. Moreover, we are also presenting some results that show the tradeoff relation between the PPP a, and some other parameters like ISP broker profits and demand blocking ratios. In DP, we are expecting to deliver an efficient resource utilization model that provides higher ISP profits and lower blocking ratios. In addition, we are expecting that our DP model can provide an efficient routing scheme. Comparing with the CS FAFA, our model can also deliver a wider room for the tradeoff relation between the demand blocking ratios, and the resulting ISP broker profits. Based on the above assumptions, we derive the following analysis:

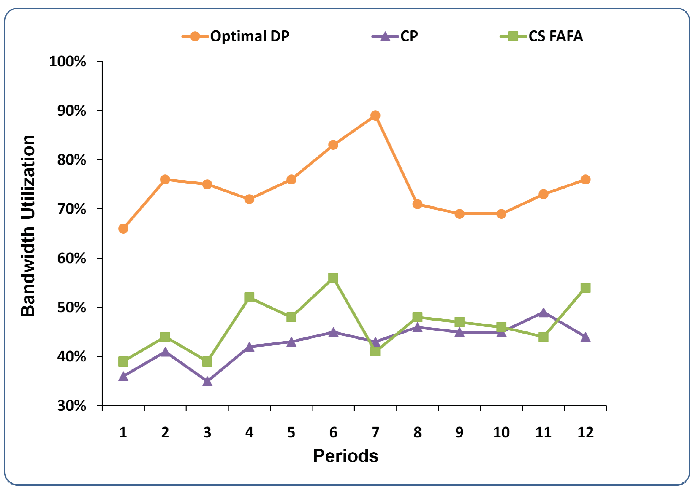


Figure 3.1 Bandwidth Utilization U_b

Figure 3.1 shows the percentage of bandwidth utilization (usage patterns) to the allocation time periods. In this figure, the DP scheme provides an average utilization of 75% of the networks' bandwidth resources through the periods 1 to 12. The CS FAFA used an average of 45%, while it is around the average of 40% for the CP case. Figure 3.2 shows the demand blocking ratios to the allocation time periods. Similarly, the DP provided the lowest blocking ratios, followed by the CP providing an average scheme comparing with the CS FAFA that shows the highest blocking ratios. In this figure, we are showing the demand blocking ratio constraint B_K^c, which we assumed to be 15% on average.

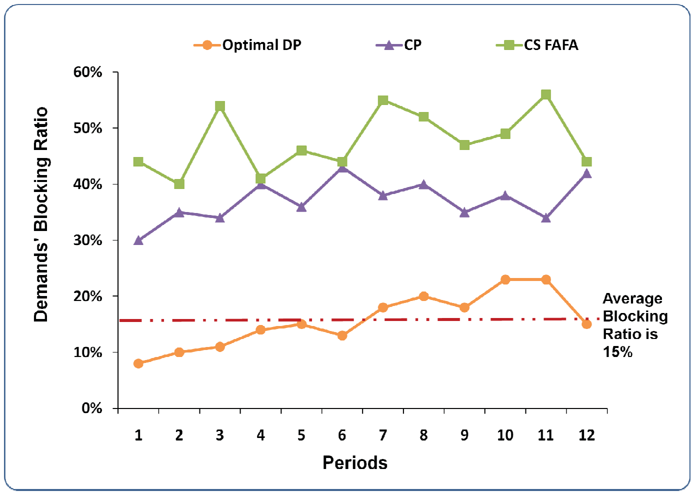


Figure 3.2 Demands' Blocking Ratio B_K

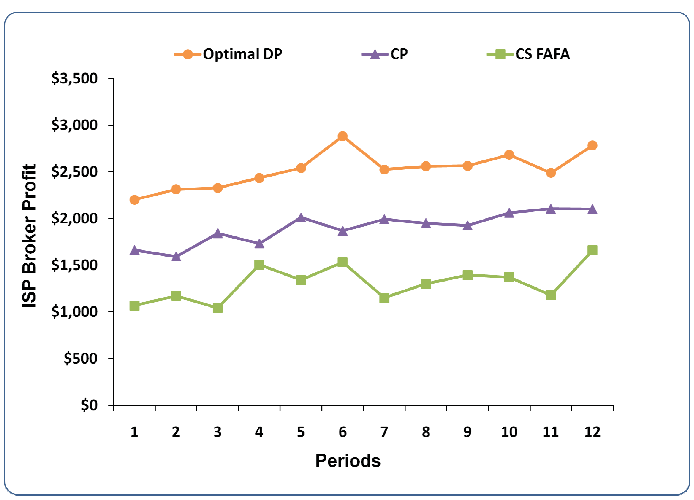


Figure 3.3 ISP Broker Total Profit P_B

Figure 3.3 shows the resulting ISP broker total profits to the allocation time periods. In this figure, it is shown that the DP provides the highest profits, followed by the CP and lastly the CS FAFA providing the lowest rates. Figure 3.4 shows the resulting ISP broker profit-units with respect to the allocation time periods. Clearly, we can remark that the highest profit-units are always provided by the DP scheme, giving an average of \$4.5 profit-unit compared with \$2.5 and \$1.5 provided by the CP and the CS FAFA, respectively. Accordingly, the DP provides an average of 3 times to what the CS FAFA provides.

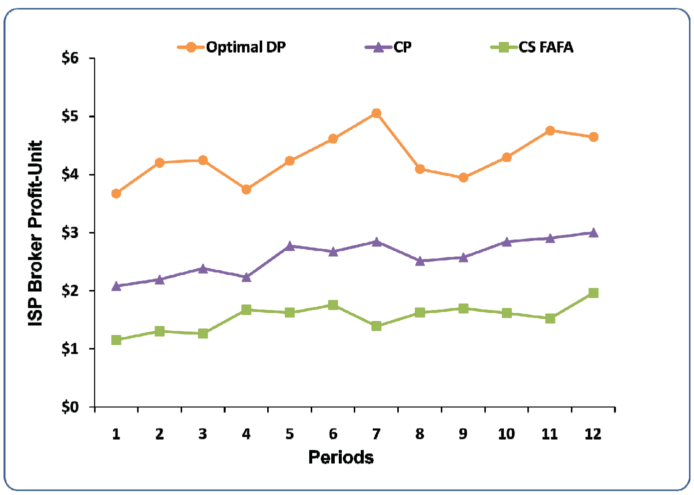


Figure 3.4 ISP Broker Profit-Unit P_{B}^{u}

Analyzing the previous figures, we can clearly notice the following: 1) The DP scheme used the largest amount of bandwidth resources, and at the same time, provided the highest profits and the best satisfaction rates represented by the lowest blocking ratios. 2) The CP scheme used the lowest amount of resources, but on the contrary, it provided higher profits and better satisfaction rates compared to that provided by the CS FAFA. 3) The CS FAFA scheme used more resources compared to that used by the CP, but still, it provided the lowest profits along with the highest blocking ratios. Consequently, we can conclude that the conducted experiments showed that the DP provides the best utilization rates of bandwidth resources. Figures 3.5, and 3.6 show results of the routing efficiency of the different partitioning schemes. In Figure 3.5, results represent the average number of hops per path used to route the admitted connection demands at the different time periods from 1 to 12, in which we can notice that the DP used the lowest average number of hops, which reflects the routing scheme efficiency. Figure 3.6 gives the number of links used to construct the routing paths for the admitted connection demands at each period. The conducted results show that the DP used the minimum number of network links, despite the fact that it served around 55% more connections compared with that served by the CS FAFA, and 30% more to what the CP served. Additionally, we can also state that experiments showed that the DP provides the most efficient routing scheme.

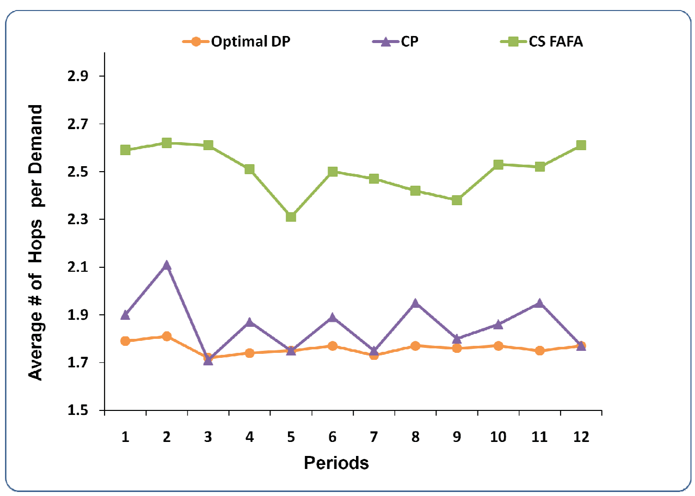


Figure 3.5 Average Number of Used Hops

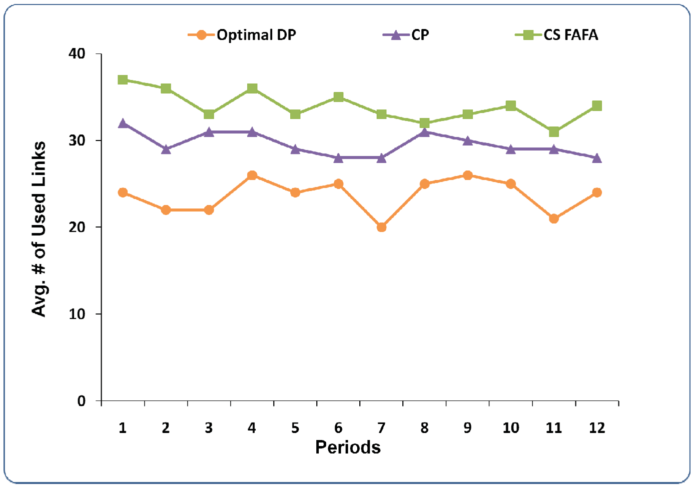
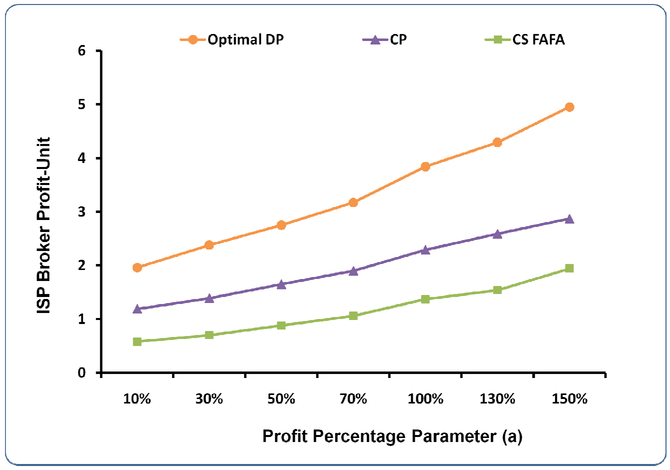
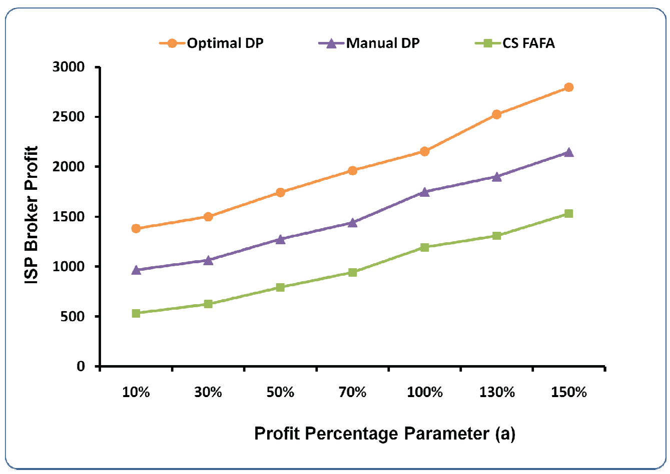


Figure 3.6 Number of Used Links

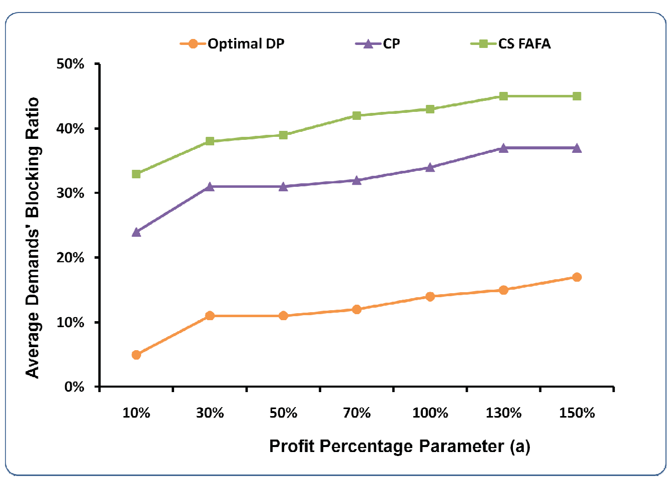
Figures 3.7.a, 3.7.b, and 3.7.c present the relationship between the *Profit Percentage Parameter* (PPP) a, Equation (3.14), and the corresponding ISP broker total profit, ISP broker profit-unit, and the demand blocking ratios, respectively. Hence, we studied the effect of different values of a ($a = 0.1, 0.3, 0.5, 0.7, 1.0, 1.3, 1.5$), and presented the results in terms of the three partitioning scenarios. As expected, results show that the DP provided the best outcomes in terms of both profits and blocking ratios. On the contrary, the CS FAFA provided the worst outcomes which reflects the weakness of the deployed resource allocation scheme. However, in the three figures, we can clearly notice a direct relationship between the PPP a and the studied performance metrics, where as the value of parameter a increases, both profits and blocking ratio increase. Accordingly, based on the demand blocking ratio constraint B_K^c or the ISP broker profit objective we can find the optimal PPP a.



(a) ISP Broker Profit vs. PPP a

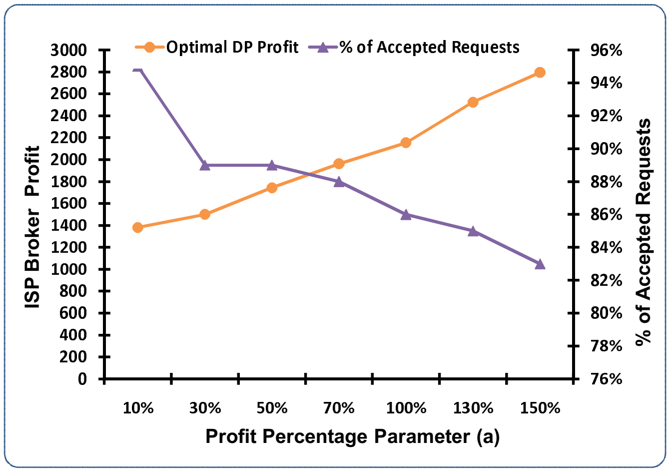


(b) ISP Broker Profit-Unit vs. PPP a

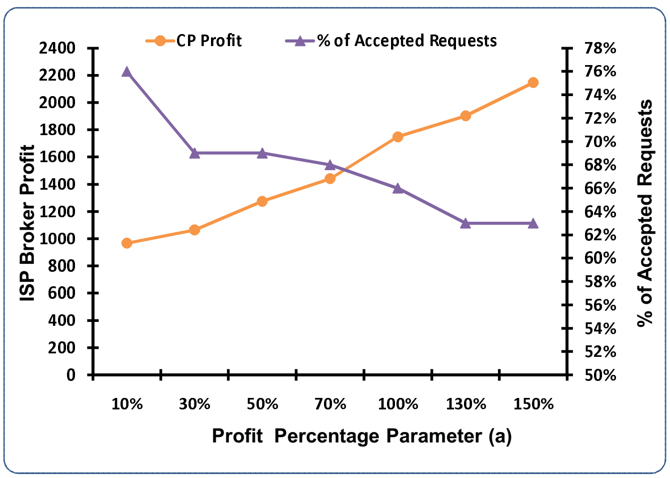


(c) Blocking Ratio vs. PPP a

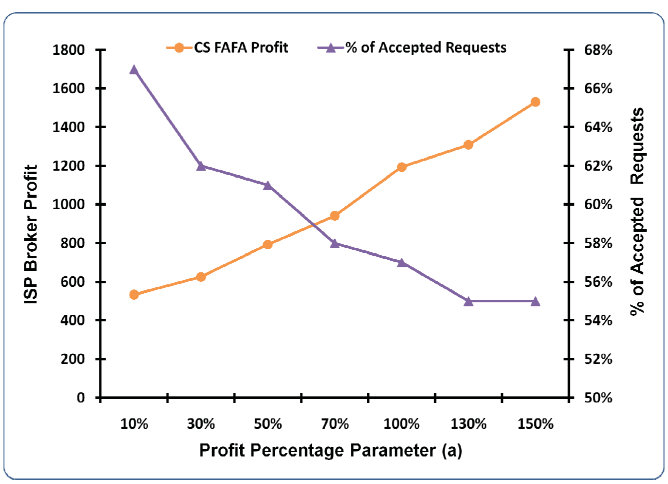
Figure 3.7 Relationship with the PPP



(a) Tradeoff in DP



(b) Tradeoff in CP



(c) Tradeoff in CS FAFA

Figure 3.8 Tradeoff with the PPP

To do so, for the given different values of a, in Figure 3.8, we present the tradeoff relationship between the ISP broker profit and the demand blocking ratios. Consequently, for a PPP of $(a = 1.3)$ we have the following: 1) With the DP, results in Figure 3.8.a show a \$2500 of total profit, and a 15% of demand blocking ratio. 2) With the CP, results in Figure 3.8.b show a \$1750 of total profit, and a 35% of demand blocking ratio. 3) With the CS FAFA, results in Figure 3.8.c show a \$1250 of total profit, and a 43% of demand blocking ratio.

3.7 Related Work

The CS model has been proposed in prior research to resolve the issue of autonomic resource management, by enabling the network users to self-manage, self-control, self-heal, and self-protect their network resources Lai *et al.* (1998); Mark *et al.* (2000). In their work, authors claimed that CS can provide a solution to the challenging management loads at the ISPs side, and deliver a satisfying resource utilization rates. Although it provides good utilization rates, but on the contrary it proved that it can lead to SLA violations. In Lai *et al.* (1998), the authors presented a comparison between both CS and CP models, where they show that the CP model can provide better QoS satisfaction rates, but on the other hand it provides less utilization. Same in Haung and Ho (2002); Cheng *et al.* (2005), authors proved that the CP model can provide higher QoS compliance rates compared to that of the CS model.

To overcome the resource utilization problem of the CP, authors in Borst and D.Mitra (1998), Blake *et al.* (1998), and Bouillte *et al.* (2002) proposed a hybrid approach that takes the advantages and overcomes the problems of both CS and CP models. Such hybrid approach is known by the Virtual Partitioning (VP) approach. Depending on the actual network traffic load, VP behaves either as CS or CP. Accordingly, it behaves as a CS at the light traffic case, while it is a CP at the extreme one Cheng *et al.* (2006b). The VP allows a heuristic resource sharing between the underloaded and the overloaded links in order to provide better resource utilization rates. Moreover, in the wireless networks domain, other authors proposed using the same hybrid VP approach. Where in Yao *et al.* (2004), an approximate analytical formulation of the VP is proposed to handle multiclass services with grand channel in a cellular system. In this work, the authors investigated a resource allocation model with preemptions for the provided service classes, to do so, they proposed using a balancing scheme that combines the *open sharing* and the *static allocation* properties of the CS and CP models, respectively. Accordingly, in heavy-load situations, resources are allocated the same way in CP, but underutilized resources are borrowed from the underloaded classes to the overloaded ones. While in the light-loaded situations, overloaded classes can use the nominal resources of all other classes. Similarly, in El-Kadi *et al.* (2002), Malla *et al.* (2003) the authors proposed a fair resource allocation protocol for multimedia wireless networks, which uses a combination of bandwidth reservation and bandwidth sharing models to provide higher utilization rates and QoS guarantees.

Such a hybrid approach sounds good, but the problem here is that the lender links (originally under-loaded) have no guarantees that they can return their resources back when they are needed. This encourages malicious over-loading. What is more, the resource sharing scheme in VP attempts a static design, in which, a pre-defined static configurations for the resource sharing process is applied at all possible traffic load conditions.

To solve this, the Bandwidth Borrowing (BR) technique is proposed to automate the resource sharing process Sup *et al.* (2005), and provide a solution for the static load-configurations attempted by the VP approach Farha and Leon-Garcia (2006), Cheng *et al.* (2006a). Accordingly, the BR technique may partially solve the bandwidth utilization problem of the CP. While yet, as long as the CP attempts a static partitioning scheme, BR cannot provide an optimal bandwidth utilization solution, where we may have many underloaded links with no overloaded ones, in such a case how the BR would work?

However, none of the above mentioned work addresses the problem of exaggeration. In fact, especially in such autonomic allocation environments, addressing the problem of exaggeration is a necessity. In our work, we are allocating the bandwidth resources based on a dynamic auction mechanism. To win the auction, auctioneers (VPN operators) should avoid any extra payments in order to submit a competing bid. This can be considered as the first motivation to prevent exaggeration. Secondly, deploying such periodic dynamic allocations will motivate the VPN operators not to exaggerate, since current allocations are valid for short times only, and so, there is no need to care about the future network changes.

It is worth to mention that in our previous works Quttoum *et al.* (2010a), Quttoum *et al.* (2010b), we proposed an efficient model that reduces the tendency of exaggeration, where we developed a threat model based on the Vickrey-Clarke-Groves (VCG) mechanism Nisan and Ronen (2007). Accordingly, exaggeration actions are reduced by charging the exaggerating users according to the inconvenience they cause to the whole system. However, the work targeted the case of autonomic resource allocation over a single network link for specific connection periods. In this work, we are dealing with the exaggeration problem in a different way, while we are targeting a full network case with multi-links and paths.

3.8 Conclusion

Based on the Linear Programming theory, we optimized the ISP profit by providing a periodic Dynamic Partitioning (DP) model for utilizing network resources. The resources are dynamically partitioned over different QoS classes through a periodic auction which can reduce the reasoning of exaggeration and maximize the ISP profit by finding the optimal set of profitable VPN connections. Such a model will eliminate the need for any borrowing technique. The advantage of our model lies in its ability of finding the optimal bandwidth division of each network link among QoS classes, and the optimal routing scheme that can guarantee the QoS commitments for the accepted VPN connections. Numerical results were conducted based on CPLEX environment where it showed that the DP model performed better than the other models. On average, the ISP profit is increased by 57%, the VPN connections blocking ratio is reduced by 64%, and network resources utilization is increased by 72%. Finally, we showed that through the proposed ILP model and the profit percentage parameter, we were able to quantify and study the tradeoff between maximizing the ISP profit and minimizing the VPN connections blocking ratio.

CHAPTER 4

DDRP: A DECENTRALIZED DYNAMIC RESOURCE PARTITIONING MODEL FOR VIRTUALIZED NETWORKS

Ahmad Nahar Quttoum[1], Abdallah Jarray[1], Hadi Otrok[2], and Zbigniew Dziong[1]

[1]Electrical Engineering Dep., École de Technologie Supérieure, Université du Québec

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2]Computer Engineering Dep., Khalifa University of Science, Technology & Research, UAE

An Article submitted to the Elsevier Journal of Network and Computer Applications in August 2011

4.1 Abstract

In this work, we study the issue of bandwidth management and profit maximization in a virtual network that is managed by a Virtual Network Operator (VNO). This VNO controls a set of network nodes that are connected through bandwidth resources leased from different Internet Service Providers (ISPs). In general, VNOs provide connectivity services in the form of Virtual Private Networks (VPNs). Resource management has a great impact on profit, especially for such VNOs where their bandwidth resources are all leased from others. The notion of *autonomic management* has emerged as a promising approach for efficient resource management. Such an approach can be deployed for different options: (1) To authorize the network users to self-manage their bandwidth resources, e.g., the models of Complete Sharing (CS) and Complete Partitioning (CP). (2) To create a *competition* environment among the network users based on *auctions*. A VNO can manage its VPNs either in centralized or decentralized schemes. Though, a decentralized scheme may reduce the *management load* at the VNO side, and provide higher *reliability* with easier *network dimensioning*. Autonomic management can be also applied to allow a decentralized management. Consequently, we propose a *Decentralized Dynamic Resource Partitioning* (DDRP) model that attempts partitioning the bandwidth resources over different QoS classes through periodic *auctions*. Hence, allocations among the

competing VPN operators go through an *auction* scenario. The management load in DDRP is distributed over Node Agents (NAs) that are spread over all network nodes representing the leasing ISPs. This creates another form of competition between the NAs to win the VPN connections, which allows a second *auction* scenario. Under such distributed interaction and the two auction scenarios, the allocation problem is modeled through a *double-auction* mechanism. This can resist exaggeration, provide *better resource utilization*, and *higher profits*. Based on the *Linear Programming theory*, we solved our model. The numerical results show the gain in performance when compared with other centralized schemes.

4.2 Introduction

Virtual Network Operators (VNOs) are providers of telecommunication services that own network nodes but use bandwidth resources that are leased from different Internet Service Providers (ISPs), Carapinha and Jimenez (2009). These ISPs do not have any management role in such networks as they are only resource suppliers for the VNOs. A VNO can be considered as a *virtual ISP* that provides connectivity services to a wide variety of customers. The most common form of such services are Virtual Private Networks (VPNs). Resource management has a great impact on profit, especially in such networks where the resources are all leased from other providers. Hence, to attain the desired profit objectives, VNOs should employ efficient resource management models. In the context of VPNs, resource management can be quite challenging to the VNOs, where VPN operators may ask for varying service requirements with different Quality of Service (QoS) levels. Connectivity services in VPNs are provided through predefined Service Level Agreements (SLAs) between the service provider - the VNO in the addressed system - and the VPN operator. Using the service provider's network infrastructure, VPN operators can provide secure and reliable connections through their virtual networks. The growing demands for such VPN services with varying QoS requirements necessitate the VNOs to efficiently utilize their leased bandwidth resources. Relying on the current management models that attempt direct interactions with the VNOs is not satisfying anymore, as they have many limitations that can be summarized as follows: (1) These models increase the management operation expenses. (2) They provide slow response times. Such limitations can lead to high rates of users' dissatisfaction. Complete Sharing (CS), Lai *et al.* (1998); Mark *et al.* (2000), and Complete Partitioning (CP), Lai *et al.* (1998); Haung and Ho (2002); Cheng *et al.* (2005), models are proposed in the literature to cope with similar management problems by creating a framework for *autonomic management*. In CS, network resources are shared over all classes without any division, while in CP, resources are statically divided among the QoS classes (or the network links), where each class (or link) uses its own allocated resources. Such approaches can be used to automate the resources allocation but at the same time they create several problems that can be summarized as follows:(1) In the CS model, resources are allocated in a First Ask First Allocate (FAFA) scheme. Also, there is no bandwidth divisions between the allocated connection classes, thus one QoS class can overwhelm all other classes in a way that reduces the ISP profit, and poses high blocking ratios. (2) In CP, resources can be under-utilized, which wastes the network resources and reduces the ISPs profits. (3) Due to the long-term SLA agreement, VPN operators might exaggerate about their required resources to guarantee their QoS in spite of unpredicted variations in the network state. Such a behavior can affect the ISP profit, and increase the blocking ratios. To overcome the problem of under-utilization in CP, the Bandwidth Borrowing (BR) technique has been proposed in Sup *et al.* (2005); Cheng *et al.* (2006a), to enable the ISPs to provide better resource utilization. According to the dynamicity of the traffic, the BR technique, and through assigned broker entities, allows the ISP to borrow the extra resources of the underloaded classes (or links) and reallocate them to the overloaded ones. However, due to the fact that such BR technique is originally built over a model that attempts a *static partitioning* scheme, BR can not guarantee an optimal utilization for the network resources. Simply, consider the case of having many underloaded links with no overloaded ones! How the BR technique can help? In such a case, the extra resources of the underloaded links will be wasted. Instead, these resources could be utilized somewhere else, i.e., the network provider can allocate them to other new demands or other networks.

In Quttoum *et al.* (2010a,b), to reduce the reasoning of exaggeration, a threat model is proposed to penalize the exaggerating users according to the inconvenience they cause to all other users in the network. Although this model provided good results in terms of exaggeration reduction, but for large networks with huge number of users, such model may impose additional computational overhead.

To solve the aforementioned limitations of the *static partitioning* models and the problem of *exaggeration*, we propose a *Dynamic Resource Partitioning* (DRP) model that attempts *dynamic partitioning* of the network resources among the provided QoS classes. Such DRP can be deployed in both, a *centralized* and a *decentralized* schemes. However, by employing the concept of *autonomic management* to create a competition environment among the network players, the latter can reduce the management load at the VNO side, and deliver higher reliability with easier network dimensioning.

In the addressed allocation problem, deploying the *centralized* DRP scheme means that the VNO is responsible for the management tasks in the whole network. This includes: (1) receiving the allocation demands from the bidding VPN operators, (2) finding the set of candidate paths for each source-destination couple, (3) dividing the ISPs links' resources among the provided QoS classes, (4) calculating the selling price thresholds for each link over the network (per QoS class), (5) routing the initially-accepted demands through the network links. On the contrary, under the *Decentralized Dynamic Resource Partitioning* (DDRP) scheme, the *management responsibilities* are less centralized, where they are distributed over special agents that are spread over all network nodes, as depicted in Figure 4.1. In DDRP, the network nodes are managed by Node Agents (NAs), where each NA represents the ISPs who provide the bandwidth resources to the assigned node. Consequently, each node is managed by NA that is responsible for leasing the bandwidth resources from the ISPs, and for any connection passing through the assigned node. All these NAs are connected to Management Agent (MA) entity that represents a broker for the VNO. Each NA, based on the expected traffic to pass through, is responsible for: (1) the division of the node's outgoing links' resources among the provided QoS classes, (2) the calculation of the desired price thresholds according to the VNO profit objectives, (3) broadcasting these metrics to the MA, and all other NAs. Thus, another competition environment is expected between the NAs. Indeed, in order to offer competing selling price thresholds to win the bidding VPN connections, NAs should search for those ISPs who lease their bandwidth resources by the lowest possible prices. In this context, the MA entity will be responsible for the rest light management tasks in the network, which includes the mapping issues between the resource sellers (NAs) and buyers (VPN operators). With the objective of maximizing the VNO profit, this MA is modeled to play an auction driven mechanism for resource allocation. Clearly, such distributed interaction and the double role of the MA entity - as resource buyer to the NAs and a seller to the VPN operators - introduces a *double-auction* framework. Consequently, this will create a form of competition between the NAs from one side, and between the bidding VPN operators from the other side. In this context, the NAs will be motivated to ask for competing selling-prices to win the received bids, while the VPN operators will be competing to reveal their highest possible bids. This makes lower price thresholds eligible, along with higher buying bids. Hence, both acceptance ratios, and the final profit rates can increase. Also, this can contribute in reducing the tendency of exaggeration, and resisting the monopolism actions.

To reduce the tendency of exaggeration without imposing any additional computational overhead, in DDRP, we propose allocating the resources based on *periodic auctions*. These periods may vary in length from few hours to several days, depending on the services provided. In such allocation scheme, the VPN operators are not expected to exaggerate their bandwidth requirements since they know that such allocations are only valid for the given period, which means that they do not have to care about future changes in the network state. Moreover, the double-auction mechanism represents another methodology to reduce the exaggeration through the internal competitions among the network players, both NAs and VPN operators.

Consequently, our contribution is a model that is able to:

- deploy a higher level of *autonomic management*, where network nodes will have higher responsibilities in self-managing their resources. This reduces the management load at the VNO side;

- provide better resource utilization than that provided by the centralized models, where decentralization allows deploying a *double-auction* scenario that crates competition among the network players;
- increase the satisfaction rates of the VPN operators, where bandwidth resources are allocated based on a *dynamic partitioning* model. This enables considering the online changes of the traffic demands;
- maximize the network profit through a *two-sided* competition environment;
- reduce the motivations of exaggeration in two ways, first through the *periodic* allocations, and second through the double-auction scenario, both without any additional computation eddenoverhead.

This model is solved through Linear Programming (LP) in a hierarchical manner. Section 4.4.3 shows the NAs' model, which is solved by each NA in a decentralized way. In contrast, Section 4.4.2.3 shows the central model that is solved by the VNO. Thus, results of the decentralized functions provide the input of the central one.

The proposed LPs provide:

- the optimal set of the profitable VPN demands;
- the optimal bandwidth division of each network link among different QoS classes;
- the optimal routing scheme of the accepted VPN demands, with respect to their QoS specifications.

Compared to other centralized models like CP, and the heuristic model of Quttoum *et al.* (2010a), Quttoum *et al.* (2010b), the numerical results show that the DDRP model performed better in terms of profit, blocking ratios, and network resource utilization.

The rest of the paper is organized as follows: Section 4.3 presents the problem statement. Section 4.4 illustrates our decentralized dynamic resource partitioning model and the proposed selection algorithm. Section 4.5 lists the proposed performance metrics, followed by the computational results in Section 4.6. In Section 4.7, we present the related work. Finally, Section 4.8 concludes the paper.

4.3 Problem Statement

The *Complete Sharing* (CS) model proposes an autonomic management framework, where the network resources are shared between all VPN operators, so each of them can use the available bandwidth at any given moment. With no link bandwidth division, in the CS approach, all QoS classes can share the network resources without discrimination. Through this, the model is supposed to ensure the delivery of services according to predefined SLAs. From the ISP prospective, giving the VPN operators such a privilege can provide better utilization of the network resources, assuming that they know their changing requirements better and accordingly utilize the available resources based on their actual needs. Naturally, as long as the VPN operators behave according to their revealed usage declarations, deploying such an approach can increase bandwidth utilization due to the use of statistical multiplexing that results in better satisfaction rates and higher profits at the same time. However, in reality, and as long as there is no link bandwidth division, VPN operators may deviate from their declarations and one class may overwhelm all other classes, which creates several problems like SLAs violations, high blocking ratios, and low profit rates.

In the *Complete Partitioning* (CP) model, the resources are partitioned among the provided service classes in a *static* way, based on a one-time bandwidth partition for all network links. With such static scheme, each class is allowed to exclusively use its portion of the provided resources. This can solve the SLAs violation problem, but on the other hand lead to low bandwidth utilization as resources might be under-utilized.

Exaggeration is considered as a common drawback of both CS and CP, where users might exaggerate their requirements in order to cope with any sudden or unpredictable future changes in the network state and link conditions, so they tend to keep a spare amount of resources that enables them to overcome and cope with such situations. This is due to the long term SLAs that do not allow users to change their resource requirements according to their needs.

To overcome the problem of resource under-utilization in CP, the *Bandwidth Borrowing* (BR) Cheng *et al.* (2006a) technique has been proposed. In the BR, the resource are borrowed from the underloaded classes (or links) and then allocated to the overloaded ones. Accordingly, the BR scheme can provide better utilization of the network bandwidth resources, and better SLAs guarantees. However, knowing that the BR technique was originally built over a model that attempts a *static partitioning* scheme, resources will not be utilized in an *optimal* way. Consider the case of having underloaded classes (or links) without any overloaded ones, in such a case the network resources will be wasted and the providers' profits will be reduced. Instead, with better resource utilization models, the network provider can exploit such extra resources to serve other demands. Consequently, the question arises *how can we solve the issue of non-optimal utilization resulting from such static partitioning scheme?*

To solve the problem of exaggeration, and based on the Vickrey-Clarke-Groves (VCG) mechanism, a threat model has been proposed in Quttoum *et al.* (2010b) to reduce the tendency for exaggeration and motivate truth-telling of the required resources. This approach provided good results, but may require additional computational overhead if applied for large networks. Therefore, one can ask: *is there any other way to reduce exaggeration rather than VCG?*

Although all of the aforementioned work deployed the concept of *autonomic management* to reduce the management load at the VNOs side, yet they are still responsible for certain *high level* management tasks. While this can be considered as a light load, in large networks with large number of users, such load can be intense. So, *how to reduce such management load?* and *how to employ such autonomic management to create a form of competition?* These questions and other mentioned before should find answers in the future management models.

4.4 The Decentralized Dynamic Resource Partitioning Model

In this section, we present our *Decentralized Dynamic Resource Partitioning* (DDRP) model that overcomes the problems of the aforementioned models, and provides higher profits and lower blocking ratios. In DDRP, as depicted in Figure 4.1, the resource allocation process is distributed through a competing NAs that are distributed over the network nodes. These NAs are connected to a central MA that represents the VNO and manages NAs from the higher level.

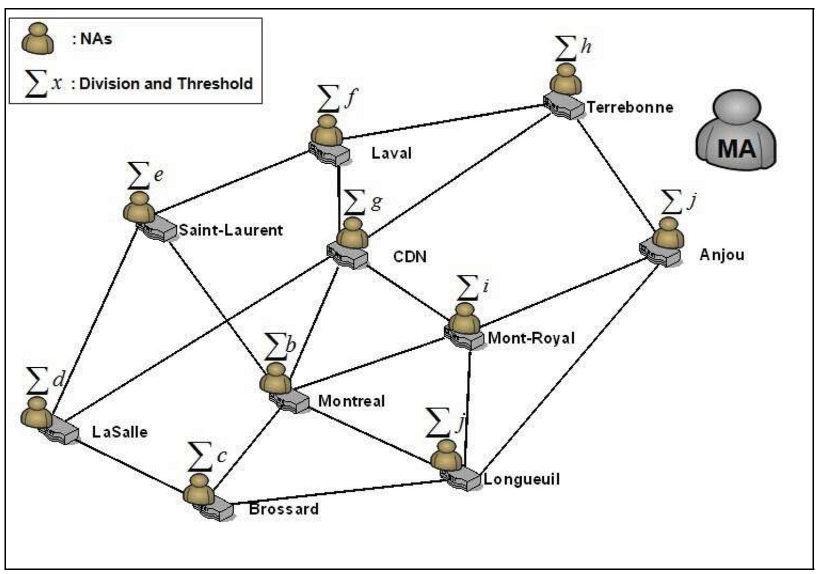


Figure 4.1 The decentralized management scheme

Each NA is responsible of its outgoing links in terms of:

- leasing the bandwidth resources form the different ISPs;
- the periodic bandwidth divisions of the links' resources among the candidate connections to pass through (division based on the QoS class);

- the calculations of the corresponding selling price threshold for each QoS class per link;
- broadcasting these division and threshold information to the MA and all other NAs.

In this context, the two main management tasks (the division, and the threshold calculations) are carried by the NAs.

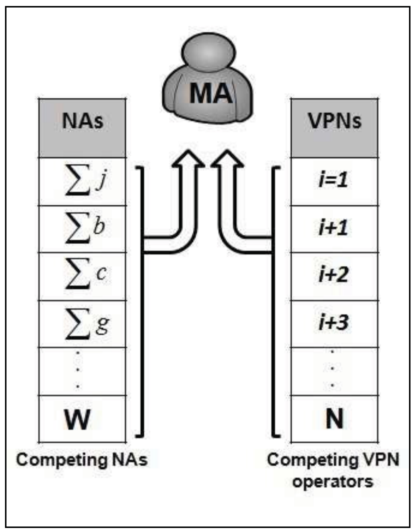


Figure 4.2 The double-auction competition scheme

Hence, as depicted in Figure 4.2, the MA responsibilities will be limited to:

- receiving the incoming bids from the competing VPN operators (demand matrix);
- receiving the links' division decisions from the competing NAs, and their corresponding selling price thresholds;
- initial acceptance/rejection of the offered bids based on the revealed information from both: NAs and VPN operators;
- routing the accepted connection demands over the winning NAs' resources.

Intuitively, in such a scenario, the MA will play a double role of being a resource buyer from the NAs and a resource seller to the VPN operators. This creates a form of competition between the NAs from one side, and between the bidding VPN operators from the other side. Accordingly, both parties will have internal competition that motivates them to reveal competing offers in order to win the auction. This can guarantee revealing lower threshold prices (NAs), and higher buying bids (VPN operators). Consequently, the acceptance ratios will be increased, and so the final profit rates. Moreover, such competing scenario can reduce the tendency of exaggeration, and resist monopolism actions.

4.4.1 DDRP Allocation Algorithm

In DDRP, resources are allocated based on two consecutive *dynamic auctions* that are held in a periodic manner, one for the selection of the VPN operators and the other for selection of the NAs. Through such competing environments, the DDRP selects the best set from the VPN operators and the NAs, and also helps in reducing the motivations for exaggeration. VPN operators' requests are represented in terms of *connection demands*, where the NAs prices to serve these connection demands represent the *tariff* for their resources. Here the term connection k is defined as a secure network tunnel that is layered on top of a public network, such tunnel has some QoS parameters (i.e. bandwidth capacity, maximum number of hops), and is used to send data from a source node s_k to another destination node d_k. The resource tariff for each NA is calculated based on the link's original cost-unit of the allocated resources, multiplied by certain profit parameter defined according to the desired profit objective of the NA. The process of selecting the winning set of VPN connections for resource allocations follows an optimal selection algorithm that is presented in Algorithm 4.1.

Consequently, at each auction period t, the MA collects the new connection demands received in the period of $(t-1, t]$ and form a new demand matrix K. Firstly, it broadcasts this matrix K to all NAs in the whole network, where each NA solves an ILP to measure the adequate QoS class divisions of its outgoing links, and calculate their corresponding price thresholds.

Once defined, each NA broadcasts its information to the MA and all other NAs. Based on that, and for each demand of the matrix K, the MA checks the offered bid and compares it with the corresponding lowest price threshold among the demand's candidate paths. For these demands that pass the previous step, the MA solves an Integer Linear Program (ILP) that is developed to choose the optimal set of VPN operators' demands and map it through the network resources in a way that provides optimal profit and bandwidth utilization rates.

Hence, comparing with the *centralized* models presented before, the proposed *decentralized* model reduces the management load at the VNO side (Algorithm 4.1: steps 2.1, 2.2, and 2.3), and creates a form of competition between the network's NAs, which positively delivers higher profits and lower blocking ratios. Moreover, comparing to the *static partitioning* models, deploying the technique of *periodic dynamic partitioning* for the network resources reduces the reasoning of exaggeration, as the bidding VPN operators will consider that such allocations stand only for the service period, and future connections will be through new auctions.

Algorithm 4.1 DDRP Selection Algorithm

Selecting the Winning VPN operators in DDRP

0: **Input:** collect the VPN operators' demands K (p_k, QoS_k, s_k, d_k);
1: At each auction round t, **the MA do**;
1.1: broadcast K to all NAs;
2: Every **NA**, and for each link of its outgoing links, solve an ILP that **do**;
2.1: choose the adequate QoS class division
2.2: calculate the threshold-prices for each QoS class;
2.3: broadcast its division and threshold decisions to the MA and all NAs;
3: Based on that, **the MA do**;
3.1: for each demand k, compare its bid p_k with the lowest path-threshold among its candidate paths;
3.2: for these demands that passed step 3.1, solve an **ILP** that finds;
3.2.1: the optimal set of VPN connections;
3.2.2: the optimal routing scheme;
4: **Output:** the VPN connections that won the resource allocation, and the MA profit collected from these connections;

4.4.2 The ILP Formulation

Through linear programming, we study the case of having a number of VPN operators, N, competing for bandwidth allocations over a network that is managed by a MA and a number of NAs, W, distributed over the network nodes. Each NA aims to sell its resources and maximize its gain. On the contrary, the MA is assumed to act on behalf of the VNO for the whole network (collection of the NAs' links), and it is responsible for dealing with the connection demand matrices and map them over the NAs according to predefined roles. Bandwidth allocation requests are modeled through connection demands, where each VPN operator i, $i \in N$, has a set of connection demands k belonging to varying sets of service classes j, $j \in J$. Logically, both VPN operators and NAs are assumed to be rational. Thus, VPN operators will aim to have their connection demands admitted with the lowest possible prices, while the NAs aims to sell their resources with the highest possible prices. On the other side, the MA aims to utilize its network bandwidth resources better, and maximizes its profit by accepting the maximum number of VPN operators while providing them with satisfactory QoS levels by competing prices (low NAs threshold prices).

Accordingly, the MA profit P function is expressed as:

$$P_{\text{MA}} = \max \left(\sum_{k=1}^{|K|} p_k - \sum_{k \in K} c_k \right) \tag{4.1}$$

where the P_{MA} is the sum of bids p_k collected from connection demands being admitted for allocation, reduced by the total cost of the bandwidth resources required to satisfy each connection demand c_k. For the MA, these c_k values are the price thresholds revealed by the NAs. Accordingly, maximizing the value of P_{MA} is achieved by maximizing the first term of the function and reducing the second term.

Hence, to maximize the first term, the MA should choose the best set of connection demands based on their bids. While choosing the set of NAs according to their selling costs, and deploying an efficient bandwidth utilization scheme to minimize the second term.

Table 4.2 QoS Classes

QoS Class	Connections Quality of Service	Bandwidth per Connection	Max. Number of Hops
5	Golden Load (<100ms Latency)	8Mbps	2 Hops
4	Excellent Load (Business Critical)	6Mbps	2 Hops
3	Controlled Load (Streaming Video)	4Mbps	3 Hops
2	Standard (IP Packet Delivery)	2Mbps	4 Hops
1	Best Effort	1Mbps	5 Hops

Choosing the best set of VPN operators' connections is typically done in accordance to their offered bids, where higher bids increases the chance for their corresponding demands to be accepted. To create a form of competition between the bidding VPN operators, we assume that the amount of bandwidth resources required to satisfy the received connection demands exceeds the available network capacity. From the NAs perspective, lower selling costs increase their chances to win more demands to be routed through their links.

Connection demands are also classified into set of different QoS classes summarized in Table 4.2, where for each auction period t, the model receives a different set of connection demands' patterns reflecting the diversity of the required services at the different auction periods. Periods can vary between day and night times, weekdays, and weekends.

4.4.2.1 Model Parameters

The studied network is defined as a set of nodes V, connected by a set of bidirectional links L, where each physical link offers certain bandwidth capacity b_l. The requests of the VPN operators N form the shape of the demands matrix K, where K represents all the connection demands submitted by the bidding VPN operators. Each demand consists of a VPN operator ID i, connection source node s_k, destination node d_k, a class of service j, $j \in J$, and a bid value p_k. Henceforth, a connection demand k of class j, consumes a bandwidth amount of b_j. Dividing the links' bandwidth capacities among the set of connection demands K and their corresponding service classes J, highly depends on the period's traffic pattern. α_j^l denotes the percentage of bandwidth capacity allocated to a connection of class j over the network link l, where $\alpha_j^l \leq b_l$.

4.4.2.2 Model Variables

To facilitate the process of measuring the MA utility presented in Equation 4.1, we refer to the admitted/rejected connections using the variable z_k, where:

$$z_k = \begin{cases} 1 & \text{if connection demand } k \text{ is admitted} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

and we also refer to the selected path that holds the considered connection using the variable x_k^π, where:

$$x_k^\pi = \begin{cases} 1 & \text{if connection demand } k \text{ uses path } \pi \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

4.4.2.3 Objective Function

According to the parameters and variables defined above, the objective function of Equation 4.1 can be reformulated as:

$$P_{\mathrm{MA}} = \sum_{k \in K} \left(p_k z_k - \sum_{\pi \in \pi_k} x_k^{\pi} \sum_{l \in \pi} c_k^l \right) \tag{4.4}$$

Indeed, the chosen paths are constructed of different links belong to different NAs, $NAs \in W$. Thus, for the MA, the original costs of the used resources are measured based on the cost-units revealed by the NAs, each for its outgoing links.

4.4.2.4 Model Constraints

For the objective function presented in Equation 4.4, we have the following constraints:

- Link Capacity: For each link l, the available bandwidth for the admitted connection demands k of class j cannot exceed the total link capacity reserved for this class.

$$\sum_{k \in K_j} \sum_{\pi \in \pi_k} \eta_l^{\pi} x_k^{\pi} b_j \leq \alpha_j^l b_l \qquad ; \, l \in L, \; j \in J \tag{4.5}$$

where the variable η_l^{π} refers to:

$$\eta_l^{\pi} = \begin{cases} 1 & \text{if path } \pi \text{ uses link } l \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

and K_j refers to the set connections belonging to QoS class j .

- Link bandwidth Division among QoS classes: The summation of the defined class divisions over any link l, $l \in L$, must not exceed the total link capacity.

$$\sum_{j \in J} \alpha_j^l = 1 \qquad ;\ \alpha_j^l \in [0, \hat{\alpha}_j^l],\ l \in L \tag{4.7}$$

where for each class j, the percentage α_j^l is bounded by the division value received from the NA managing the considered link $\hat{\alpha}_j^l$. Hence, for each link, such value represents an upper-bound for the MA's division decision.

- Acceptance Threshold-price: To be considered as a competing connection demand for the allocation process, the minimum offered bid p_k for a connection demand k of QoS class j must at least be higher than or equal to a certain threshold. The calculation of this threshold depends on the revenue objectives of each NA, which is out of the scope of this work.

$$p_k \geq \sum_{\pi \in \pi_k} x_k^{\pi} \sum_{l \in \pi} p_{th,j}^l \qquad ;k \in K_j,\ j \in J \tag{4.8}$$

- Routing Path Assignment: Only one routing path can be assigned to carry each admitted connection.

$$\sum_{\pi \in \pi_k} x_k^{\pi} \leq 1 \qquad ;\ x_k^{\pi} \in \{0, 1\},\ k \in K \tag{4.9}$$

- Linking decision variables: Finding a path for the connection demand does not guarantee final acceptance, the offered bid p_k must also satisfy the constraint represented in Equation 4.8.

$$z_k \leq \sum_{\pi \in \pi_k} x_k^{\pi} \qquad ; z_k \in \{0,1\},\ k \in K \tag{4.10}$$

- Routing Path Length: The length of the assigned path l_π for a connection demand k of QoS class j cannot exceed H_j hops, one hop represents one physical link.

$$l_\pi \leq H_j \qquad ; \pi \in \pi_k,\ j \in J \tag{4.11}$$

4.4.2.5 Model Complexity

The computational complexity of the model can be measured through the number of model variables and constraints, as follows:

- Number of variables:

$$|K| + |K| \times |\pi_{\mathrm{K}}| + |L| \times |J|$$

  which can be simplified as:

$$|K| \times |\pi_{\mathrm{K}}| + |L| \times |J|$$

- Number of constraints:

$$|L| \times |J| + 3|K| + |K| \times |\pi_{\mathrm{K}}|$$

which can be simplified as:

$$| L | \times | J | + | K | \times | \pi_{\mathrm{K}} |$$

From this, we can conclude that the number of variables equals the number of constraints, which is in the order of $O(n^2)$. Moreover, to deal with the complexity issues, if exist, we can use the technology of *Cloud Computing* that is considered as an efficient solution to deal with high combinatorial problems.

4.4.3 The Link Bandwidth Division Model

To assign the VPN operators' connection demands to the most profitable paths, the DDRP model executes the following steps:

- at each allocation period t, and in accordance to the received demand matrix and the links' capacities, each NA solves an ILP that generates a class division map. This map shows the percentages of bandwidth capacities it reserved for each QoS class over each link of its outgoing links;
- based on the NAs' division maps and their revealed price thresholds, for each demand k, the MA finds the set of candidate paths Epstein (1994) to carry its connection;
- the MA's ILP chooses the most profitable combinations of the VPN operators' demands to be carried over the network links, and consequently assign them to their routing paths.

As an input to the MA's ILP, in a decentralized way, each NA solves the following ILP that represents the NA's utility function:

$$P_{\mathrm{NA}} = \mathrm{Max} \sum_{k \in K} \left(p'_k z_k - \sum_{\pi \in \pi_k} x_k^{\pi} \sum_{l' \in \pi} c_k^{l'} \right) \tag{4.12}$$

where each NA aims to maximize its profit P_{NA} by: (1) maximizing the sum of bids to be collected from the demand connections to pass through. The term p_k' refers to the percentage of bid it would earn for its resources from the total bid of the whole connection. (2) minimizing the sum of resources' original costs by allocating (utilizing) its network resources efficiently, sure based on efficient division metrics. Such resource utilization is motivated by the competing environment between the NAs, where through the auction scenario, they will be competing to offer their links' resources by lower prices to win more connection demands. To afford this, rational NAs should keep the usage of their available resources efficient (use the lowest possible bandwidth amounts). This will enable them to offer lower prices, see Figure 4.3 for the assumed pricing scheme. The assumption of such approximate pricing curve is used to emulate the case of having limited amounts of bandwidth resources at the main providers' side. Thus, to motivate using lower amounts of resources, the providers follow an inverse relationship between the selling prices, and the residual bandwidth resources. Hence, rational VPN operators who aim to maximize their profits will try to use the lowest possible amounts of bandwidth resources to avoid higher bandwidth costs.

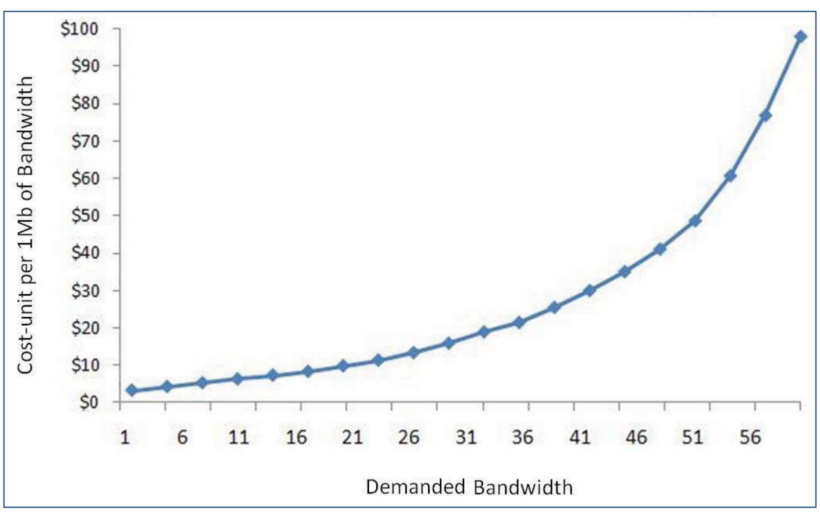


Figure 4.3 Approximate Pricing Curve

To create a form of competition, the assigned candidate paths might be shared between various connection demands having the same source-destination nodes, where the sum of bandwidth amounts over these candidate paths is less than or equal to that required to satisfy the whole connections with the same source-destination couples. In this context, it is worth to mention that the link bandwidth division among the different QoS classes depends mainly on the demands' diversity, where such diversity reflects the different VPN operators requirements from time to time. Differences may appear according to the day time (i.e. morning, afternoon, evening, and night), the weekday (i.e. working days, and weekends), or even it could be also affected by the different time-zones.

However, in this work we are presenting two different scenarios of link bandwidth division among QoS classes. In the first scenario, we are showing the case of CP, where the network resources are statically partitioned based on statistics fed to the model in a *heuristic* way. In the second scenario, the ILP model solves the whole resource allocation problem (centralized and decentralized) and automatically generates the *optimal* link bandwidth division map among QoS classes. Empirical results in Subsections 4.6.1 and 4.6.2 show comparison between the aforementioned scenarios, in terms of total profits, bandwidth utilization, demand blocking, end-to-end delays and routing schemes efficiency.

4.4.4 Benchmark Models

To judge our DDRP mode, we proposed to compare it with the CP mode that attempts a static-partitioning scheme. Moreover, we also provide a comparison with the centralized dynamic allocation model presented in Quttoum *et al.* (2010b), but with few modifications explained below. This modified centralized model is named by the *Centralized Dynamic Resource Partitioning* (CDRP) model. It is worth to mention that we did not choose to compare DDRP with the CS model, since we believe that this does not provide a fair comparison. This follows from the fact that the CS attempts a FAFA admission control scheme for resource allocation, which does not consider the profit objectives we are considering in this work.

4.4.4.1 The CDRP Model

In the modified centralized model:(1) the allocation process is applied in a dynamic-periodic manner, (2) the links bandwidth resources are divided between the QoS classes through a central entity, and (3) the model is solved through linear programming. So, the dynamic class division process will take place in a periodic auction manner, exactly the same as in DDRP, while the division and threshold calculations are done by the VNO (represented by the MA). Hence, the model will follow the same allocation process that is presented in Algorithm 4.1, except steps 3, 4, and 5 which are processed by the centralized entity instead of the NAs.

4.4.4.2 The CP Model

In CP, the model follows a static division map for the network links. Division takes place *only once* at the start of the network planning period, and follows a heuristic scheme based on the shortest path algorithm. For fair comparison, we limit the differences between the compared models to the allocation process and division schemes. Thus, we solved the CP model through linear programming as well. The allocation process of the CP is depicted in Algorithm 4.3.

Algorithm 4.3 CP Selection Algorithm

Selecting the Winning VPN operators in CP

1: **Input:** Collect the VPN connections K (p_k, QoS_k, s_k, d_k) received through time $(t-1,t]$;
2: At each auction round t, the VNO **do**;
2.1: Based on the shortest path algorithm, find the adequate QoS class divisions for the network links;
2.2: Based on 2.1, formulate the problem as an ILP;
2.3: Solve the ILP, and find the optimal set of VPN connections;
3: **Output:** the VPN connections that won the resource allocations, and the profits collected from these VPN connections;

4.5 Performance Metrics

To evaluate our DDRP model we make comparison with CDRP and CP models discussed in the previous section. For each model we are calculating the following metrics:

4.5.1 The Network Total Profit

The profit P is measured based on the bids collected from the admitted connection demands, reduced by the cost of the network links. Both bids and costs are expressed in terms of $\$X$.

4.5.2 The Network Profit-Unit

The profit-unit value represents the gain collected per $1Mb$ of bandwidth, which is measured as the ratio between the total profit measured in Subsection 4.5.1, to the amount of bandwidth of all links:

$$P^u = \frac{P}{\sum\limits_{l \in L} b_l} \tag{4.13}$$

4.5.3 Demands' Blocking Ratio

Demands' blocking ratio represents the VPN operators' satisfaction rates, and is measured as the ratio between the number of admitted connection demands to the number of all demands participated in the allocation auction:

$$B_K = \frac{\sum\limits_{k \in K} z_k}{|K|} \tag{4.14}$$

4.5.4 Bandwidth Utilization

Bandwidth Utilization is measured as the ratio between the used and the total bandwidth amounts:

$$U_b = \frac{\sum_{j \in J} \sum_{k \in K_j} b_j z_k}{\sum_{l \in L} b_l} \tag{4.15}$$

4.5.5 Routing Scheme Efficiency

To evaluate the routing scheme efficiency of the above mentioned scenarios, we proposed measuring the following parameters:

4.5.5.1 End-to-End Delay

The average end-to-end delay for admitted connection demands is measured through the average number of *hops* used to form the routing paths:

$$H = \frac{\sum_{k \in K} \sum_{\pi \in \pi_k} x_k^\pi l_\pi}{\sum_{k \in K} z_k} \tag{4.16}$$

4.5.5.2 Number of Used Network Links

The number of *links*, L^u, that is used to form all paths in the assigned period of time.

4.6 Computational Experiments

To illustrate the efficiency of this work, under the CPLEX environment, we illustrate an example of a network that consists of 6 nodes, connected by 14 bidirectional links, for which we assume that each link holds a $60Mb$ of bandwidth resources. For this network, we assume receiving 100 VPN operators' connection demands per period of time, distributed as 20% for QoS_1, 30% for QoS_2, 15% for QoS_3, 15% for QoS_4, and 20% for QoS_5. In this scenario, we assume dividing the week planning time into 12 classes, each stands for 6 hours of time Thompson *et al.* (1997). This represents a new allocation auction every 6 hours. However, the period between such auctions may vary according to the offered services and their durations.

Demands with different service durations can be divided into groups, each follows certain period length. In this example, we are assuming selling certain service packages that stand for 6 hours of time (e.g. online games, video conferencing, etc). Table 4.4 below shows the periods' division map. Through this network scenario, we will study the behavior of the above mentioned three models (DDRP, CDRP, and the CP).

Table 4.4 Periods' Division Map

Period Index	Corresponding Time Period
1	Monday to Friday: Morning
2	Monday to Friday: Afternoon
3	Monday to Friday: Evening
4	Monday to Friday: Night
5	Saturday: Morning
6	Saturday: Afternoon
7	Saturday: Evening
8	Saturday: Night
9	Sunday: Morning
10	Sunday: Afternoon
11	Sunday: Evening
12	Sunday: Night

4.6.1 Bandwidth Division Efficiency

In the studied models, different bandwidth division schemes are followed. As mentioned before, the CP model attempts a heuristic bandwidth division scheme that is handled directly by the VNO or its broker (i.e. based on the shortest path algorithm). On the contrary, such a bandwidth division process is handled in a *dynamic* way through ILPs in both CDRP and DDRP models. In CDRP, such a process is solved within a centralized entity, usually the VNO or its broker. While in DDRP, it is distributed through the NAs that are spread over the network nodes; each NA is responsible for dividing the resource of its outgoing links. For these three scenarios, Tables 4.5, 4.6, and 4.7 are showing the provided QoS classes, their acceptance ratios, their bandwidth usage amounts, and their corresponding profits at three different periods. For better comparison, in each table, we present the results of different periods that vary between weekdays, Saturdays, and Sundays.

Table 4.5 DRRP performance

QoS Class	*Week-Day: from 6 am to 12 pm*			*Saturday: from 6 pm to 12 am*			*Sunday: from 6 pm to 12 am*		
	Acceptance Ratio	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*
1	100%	30	530	100%	30	652	100%	30	579
2	90%	124	993	100%	124	1185	97%	116	1028
3	100%	96	950	94%	96	1134	100%	100	1198
4	67%	132	585	73%	132	789	73%	132	723
5	80%	200	1088	75%	200	1240	65%	176	1020

The *week-day* period in the tables shows that under the DDRP model we accepted around 88% of the connection demands, compared to 68% and 54% in the CDRP and CP models, respectively. Through the same period, the DDRP model shows utilization of 582Mb of the available bandwidth, while it is 406Mb in the CDRP, and 273Mb in CP. Resulting profits are showing that the DDRP provides a value of \$4146, representing an increase of 35% and 72% to that provided by the CDRP and CP models, respectively.

Table 4.6 CDRP performance

QoS Class	*Week-Day: from 6 am to 12 pm*			*Saturday: from 6 pm to 12 am*			*Sunday: from 6 pm to 12 am*		
	Acceptance Ratio	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*
1	100%	30	480	100%	30	602	85%	25	522
2	73%	88	866	77%	92	1065	77%	92	884
3	80%	76	758	67%	64	918	67%	60	954
4	47%	84	348	53%	96	495	47%	84	390
5	40%	128	616	45%	136	792	45%	136	724

In the *Sunday* period, we can see that the DDRP model accepted around 87% of the received demands, compared to that of 65% and 54% in CDRP and CP, respectively. In the bandwidth usage, the DDRP also shows a utilization of 554Mb of the available resources, where it is

397*Mb* in the CDRP and 291*Mb* in CP. Concerning the profit, the DDRP provides a $4548, showing an increase of 30% and 72% to that provided by the CDRP and CP, respectively. Same conclusions are noticed from records concerning the *Saturday* period.

Table 4.7 CP performance

QoS Class	*Week-Day: from 6 am to 12 pm*			*Saturday: from 6 pm to 12 am*			*Sunday: from 6 pm to 12 am*		
	Acceptance Ratio	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*	*Acceptance Ratio*	*Bandwidth Usage*	*Total Profit*
1	95%	29	478	95%	28	595	80%	23	516
2	70%	84	855	80%	98	1062	77%	96	840
3	66%	68	634	66%	68	768	73%	80	812
4	20%	36	171	20%	36	189	20%	36	171
5	20%	56	268	20%	56	336	20%	56	308

In summary the comparison the three different scenarios shows that the DDRP model always provide the best results in terms of: *acceptance ratios*, *efficient bandwidth utilization* and *final profit rates*, followed by the CDRP and lastly the CP. This reflects the efficiency of the bandwidth division scheme attempted by the DDRP.

4.6.2 General Performance and Routing Scheme Efficiency

Section 4.6.1 provided sample numerical results of the DDRP model compared to the CDRP and CP models. In this section, we are providing a wider range of comparison through the allocation periods from 1 to 12 between the aforementioned models in terms of: (1) Customers' satisfaction represented by the blocking ratios. (2) Bandwidth utilization, which represents how the bandwidth resources were used. (3) Total profits resulting from the previous two metrics. (4) Profit-units. (5) Routing efficiency in terms of the average end-to-end delay (average number of used hops), and the number of used links.

Figure 4.4 shows the percentage of the demands' blocking for different allocation periods. In this figure, the DDRP model provides an average of 86.5% acceptance ratio through the periods 1 to 12. While it is on the average of 66.75% in the CDRP, and around 57.75% in CP. Figure 4.5 presents the percentage of bandwidth utilization (usage patterns) to the allocation periods.

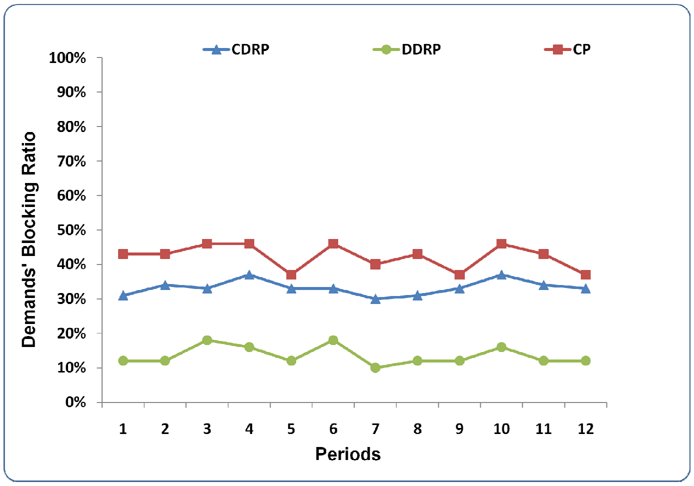


Figure 4.4 Demands' Blocking Ratio B_k

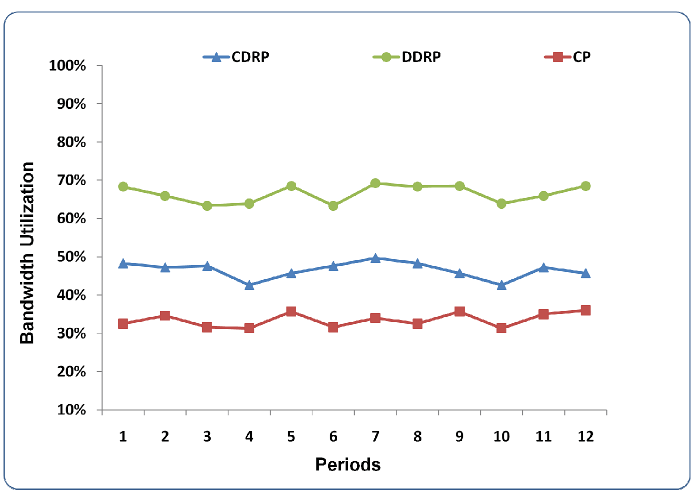


Figure 4.5 Bandwidth Utilization (usage)U_b

On average, through the periods 1 to 12, the DDRP shows a 66.5% utilization of the network bandwidth resources. On the contrary, it is 46.6% in the CDRP and 33.6% in the CP. In Figure 4.6, we are showing the resulting total profits to the allocation periods. Accordingly, it is shown that the DDRP provides the highest profits with the average of $3757, showing an average increase of 36.3% to the CDRP and 72.5% to the CP. This reflects the result of having better utilization rates of the bandwidth resources in the DDRP, which enables the model to accept more connection demands, and hence higher profits.

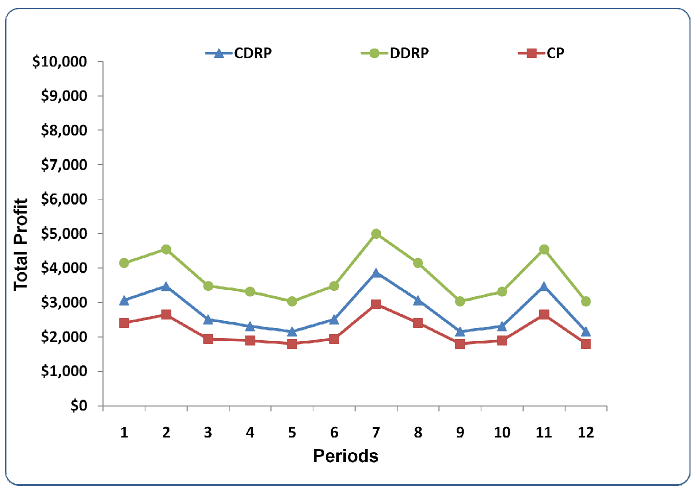


Figure 4.6 Total Profit Ratio *P*

Based on the above we can draw the following conclusions: (1) The DDRP model used the largest amount of bandwidth resources, but in an efficient way. At the same time, it provided the highest profit rates, and the best satisfaction represented by the lowest blocking ratios. (2) The CDRP model showed an average performance compared to the DDRP and CP models. (3) The CP model used less resources compared to that used by the DDRP and CDRP, provided the lowest profits, and the highest blocking ratios.

Figure 4.7 shows the resulting profit-units with respect to the allocation periods. Although, we can remark that the highest profit-units are always provided by the CP model, this does not necessarily mean higher *total profits*. Having such high blocking ratio (see Figure 4.4) in the CP case, with low bandwidth usage (Figure 4.5), and comparatively high profit (Figure 4.6) will definitely result in high profit-units. One possible explanation is that the CP model chooses only the demands with the high bids for allocation, without considering the blocking ratios, and the other management metrics. However, what concern more is the total final profit. As shown from the results, the DDRP is always leading in terms of total profits.

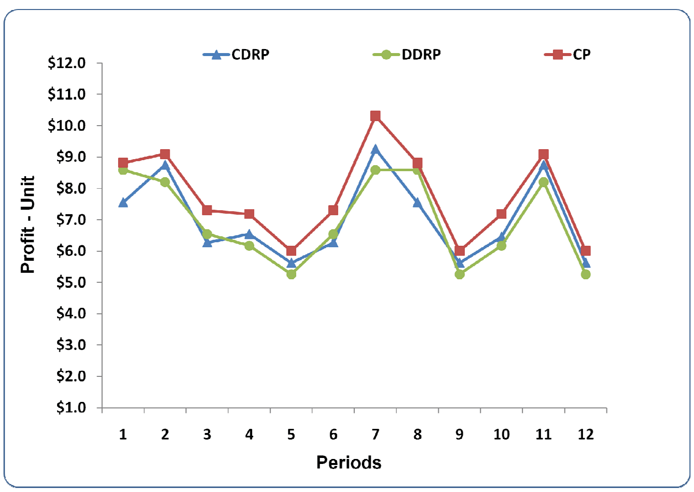


Figure 4.7 Profit-Unit

Figures 4.8, and 4.9 show the efficiency of the routing algorithm, from which we can conclude that although the DDRP model used approximately the same number of hops and links used in CP, it was able to use the available paths more efficiently and deliver higher satisfaction and profit ratios. In terms of demands' satisfaction, on average, the DDRP served 30% and 50% more connections to that served by the CDRP and CP, respectively.

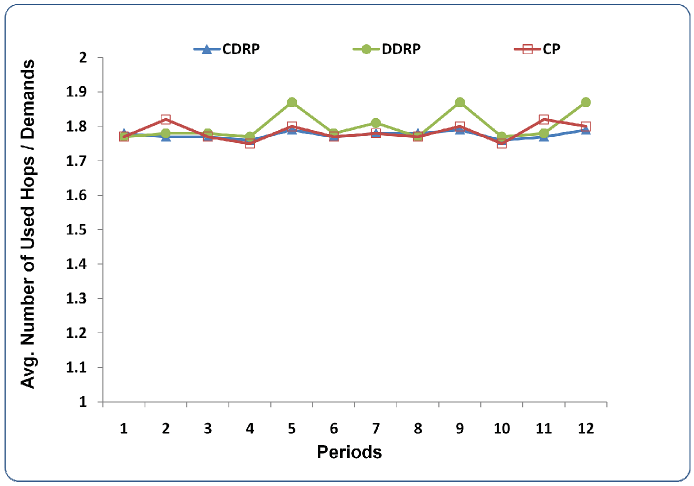


Figure 4.8 Average No. of Hops

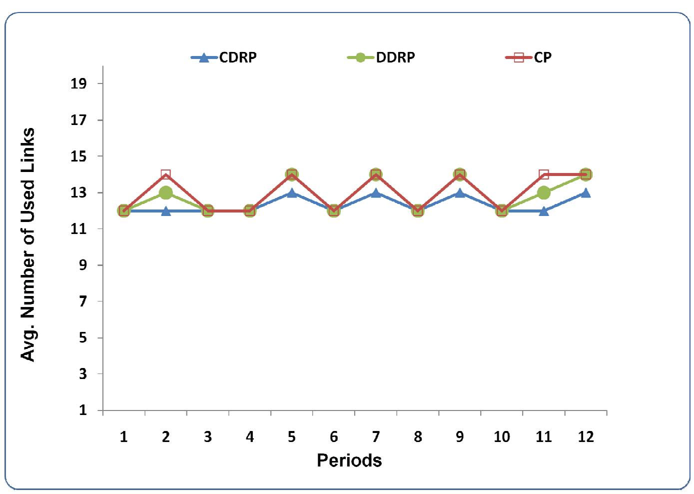


Figure 4.9 No. of Used Links

4.7 Related Work

The CS model has been proposed in prior research to resolve the issue of autonomic resource management, by enabling the network users to self-manage, self-control, self-heal, and self-protect their network resources Lai *et al.* (1998); Mark *et al.* (2000). In their work, authors claimed that CS can provide a solution to the challenging management loads at the ISPs side, and deliver a satisfying resource utilization rates. Although it provides good utilization rates, but on the contrary it proved that it can lead to SLA violations. In Lai *et al.* (1998), the authors presented a comparison between both CS and CP models, where they show that the CP model can provide better QoS satisfaction rates, but on the other hand it provides less utilization. Same in Haung and Ho (2002); Cheng *et al.* (2005), authors proved that the CP model can provide higher QoS compliance rates compared to that of the CS model.

To overcome the resource utilization problem of the CP, authors in Borst and D.Mitra (1998), Blake *et al.* (1998), Bouillte *et al.* (2002) proposed a hybrid approach that takes the advantages and overcomes the problems of both CS and CP models. Such hybrid approach is known by the Virtual Partitioning (VP) approach. Depending on the actual network traffic load, VP behaves either as CS or CP. Accordingly, it behaves as a CS at the light traffic case, while it is a CP at the extreme one Cheng *et al.* (2006b). The VP allows a heuristic resource sharing between the underloaded and the overloaded links in order to provide better resource utilization rates. Moreover, in the wireless networks domain, other authors proposed using the same hybrid VP approach. Where in Yao *et al.* (2004), an approximate analytical formulation of the VP is proposed to handle multiclass services with grand channel in a cellular system. In this work, the authors investigated a resource allocation model with preemptions for the provided service classes, to do so, they proposed using a balancing scheme that combines the *open sharing* and the *static allocation* properties of the CS and CP models, respectively. Accordingly, in heavy-load situations, resources are allocated the same way in CP, but underutilized resources are borrowed from the underloaded classes to the overloaded ones. While in the light-loaded situations, overloaded classes can use the nominal resources of all other classes. Similarly, in El-Kadi *et al.* (2002); Malla *et al.* (2003) the authors proposed a fair resource allocation protocol for multimedia wireless networks, which uses a combination of bandwidth reservation and bandwidth sharing models to provide higher utilization rates and QoS guarantees. Such a hybrid approach sounds good, but the problem here is that the lender links (originally under-loaded) have no guarantees that they can return their resources back when they are needed. This encourages malicious over-loading. What is more, the resource sharing scheme in VP attempts a static design, in which, a pre-defined static configurations for the resource sharing process is applied at all possible traffic load conditions.

To solve this, the Bandwidth Borrowing (BR) technique is proposed to automate the resource sharing process Sup *et al.* (2005), and provide a solution for the static load-configurations attempted by the VP approach Farha and Leon-Garcia (2006); Cheng *et al.* (2006a). Accordingly, the BR technique may partially solve the bandwidth utilization problem of the CP. While yet, as long as the CP attempts a static partitioning scheme, BR cannot provide an optimal bandwidth utilization solution, where we may have many underloaded links with no overloaded ones, in such a case how the BR would work?

However, none of the above mentioned work addresses the problem of exaggeration. In fact, especially in such autonomic allocation environments, addressing the problem of exaggeration is a necessity. In our work, we are allocating the bandwidth resources based on a dynamic auction mechanism. To win the auction, auctioneers (VPN operators) should avoid any extra payments in order to submit a competing bid. This can be considered as the first motivation to prevent exaggeration. Secondly, deploying such periodic dynamic allocations will motivate the VPN operators not to exaggerate, since current allocations are valid for short times only, and so, there is no need to care about the future network changes. It is worth to mention that in our previous works Quttoum *et al.* (2010a,b), we proposed an efficient model that reduces the tendency of exaggeration, where we developed a threat model based on the Vickrey-Clarke-Groves (VCG) mechanism Nisan and Ronen (2007). Accordingly, exaggeration actions are reduced by charging the exaggerating users according to the inconvenience they cause to the whole system. However, the work targeted the case of autonomic resource allocation over a single network link for specific connection periods. In this work, we are dealing with the exaggeration problem in a different way, while we are targeting a full network case with multi-links and paths.

4.8 Conclusion

Based on the Linear Programming and the double-auction mechanism, we optimized the VNO profit by providing a periodic decentralized dynamic partitioning model for utilizing the network resources. The resources are dynamically partitioned over different QoS classes through two periodic auctions, which can reduce the reasoning of exaggeration, and maximize the network profit by finding the optimal set of profitable VPN connections with price competing resources. Distributing the bandwidth division processes over the network nodes can provide more reliable management systems, and easier network dimensioning. Compared to the CDRP, deploying the decentralized model allows a double-auction competing environment. This can deliver better resource utilization, higher profits, and better satisfaction ratios. The advantage of our model lies in its ability to find the optimal bandwidth division of each network link among QoS classes, finding the optimal routing scheme that can guarantee the QoS commitments for the accepted VPN connections. Another major contribution is the reduction of the management load at the VNO side. Numerical results were conducted based on CPLEX environment, and it showed that the DDRP model performed better than the other models. On average, the profit is increased by 73%, the VPN connections blocking ratio is reduced by 68%, and network resources utilization is increased by 98% when compared to the *static partitioning* model (CP).

CONCLUSION

Network virtualization combines the concepts of *virtual partitioning* and *autonomic management*. While complete autonomic management systems have not been delivered yet, this work provides a considerable progress toward that goal. Previously, as presented in Section 1.2, network resource allocation for VPNs was mostly manual, requiring direct intervention from the ISPs. Later on, the concept of autonomic management started to be employed, but with no virtual partitioning. While these models provided better resource utilization and reduced management load at the ISPs side, still selfish actions like resource overwhelming and exaggeration were not addressed. Next, the autonomic management with virtual partitioning was deployed in order to solve the problems of exaggeration and resource overwhelming. In this case, the network bandwidth resources were virtually partitioned among the network customers in a *static* way. This solves the problem of resource overwhelming, but does not solve the problem of exaggeration that might happen at the SLA contracting stage. In addition, it results in non optimal resource utilization due to the static virtual partitioning. To overcome this, other techniques were also proposed to adapt the static resource allocations by borrowing the extra resources from the underloaded partitions, and allocate it to the overloaded ones. Although this enhances the utilization rates, but still, its outcomes are not optimal.

The work of this thesis was proposed to overcome such problems and limitations. To achieve better resource utilization, Chapter 2 presents how the concepts of *auction mechanism*, and *autonomic management* can be deployed. The idea of deploying the auction mechanism provides a smart way to choose the profitable set of customers that are able to choose the adequate amount of bandwidth resources to satisfy their needs. Derived from the *Game theory*, the VCG mechanism is applied to reduce the tendency of exaggeration that is expected to result from the autonomic management privilege. The VCG mechanism provides a powerful tool to enforce cooperation in such non-cooperative scenarios. Also, from the *Markov Decision theory*, a minimum selling price value is calculated based on the state-dependent shadow price concept. This value represents the minimum selling-threshold that is used to define whether to accept or reject the customers' offered bids. This protects against the harmful collusion actions that can be expected in the context of auction environments.

In Chapter 3, the resource allocation problem is expanded from a link problem to a full network that includes *bandwidth partitioning* among the carried QoS classes. Linear Programming theory is applied to provide the *optimal* management metrics for such a problem. Different from the exaggeration resistant model presented in Chapter 2, in this part, the resource partitioning process was executed dynamically in a *periodic* way. Such periodic scenario showed an efficient way to reduce the exaggeration motivations, without the need of the threat model. Periodicity allowed convincing the customers that exaggerating their bandwidth requirements is not beneficial any more, since comparing to the previous allocation scenarios, current allocations are only valid for short periods of time.

In Chapter 4, the work addressed the problem of resource management in networks that are managed by Virtual Network Operators (VNOs). This considered the case in which VNOs lease their network bandwidth resources from different ISPs. Compared to the previous propositions in Chapters 2 and 3, a new model is developed to solve the resource management problem in a *decentralized* way, in which, the VNO became able to distribute the management load over special entities that are employed to manage the network nodes. Namely, these entities are called Node Agents (NAs). This allowed the VNOs to employ a *higher level of autonomic management*, where not only the network customers, but also the NAs are responsible to self-manage and control their resources. Indeed, such a model deploys a two-sided autonomic management scheme that allowed a *double-auction* framework. This creates a competing environment between the NAs (resource sellers) from one side, and the customers (buyers) from the other side. Consequently, this helps in keeping the NAs motivated to use their available resources better in order to offer competing selling prices and win the customers' bids. The same applies for the network customers, where they compete to offer higher bids to win the resource allocations from the NAs. The solution is sloved hierarchically through the Linear Programming theory.

Achieved Objectives

The work presented in Chapter 2 provided better usage of the bandwidth resources, which resulted in *higher profits* and *lower blocking ratios* when compared to the FAFA bandwidth allocation algorithm,. Moreover, it delivered efficient methods to reduce the effect of the *exaggeration* and *collusion* cheating actions. Hence, this part of the work answered the following questions (presented in the objective list in Section 0.3):

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to suppress collusion actions that harm the allocation process?

In Chapter 3, the work provided new techniques for optimal bandwidth partitioning. Such techniques helped in *utilizing the network bandwidth resource better*, gain *higher profits* with *lower blocking ratios*. Compared to the previous part, this chapter provided different methodologies to answer the first two objective questions. Indeed, it reduced the *exaggeration* motivations through the *periodic auctions* instead of the threat model used in Chapter 2. Same for the resource utilization, the Linear Programming has been employed to solve an *optimal* dynamic resource partitioning model. From the list of Section 0.3, this part of the work answered the following questions:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to measure the optimal bandwidth division among QoS classes?

Thus, this part of the work The decentralized model of Chapter 4 proposed a new model based on a *higher level* of autonomic management, with a *double-auction* scenario. Through which, it provided *higher bandwidth utilization rates*, which implicitly means *higher profit and satisfaction ratios*. The double-auction allowed internal competition within the NAs and the network customers. Within NAs, this resulted in *better utilization* and *lower blocking ratios*, while it delivered *higher total profits* from the competition among the network customers. Moreover, in addition to the reduction of the *exaggeration* motivations due to the periodic allocations, in this part, the two-sided competing environment represented *another new way* to reduce exaggeration. Thus, the achievements of this part answered the following questions:

- How to provide better utilization of the network resources?
- How to motivate VPN operators not to exaggerate?
- How to measure the optimal bandwidth division among QoS classes?
- How to decentralize the resource allocation mechanism?

It is worth to mention that compared to the work of Chapter 3, and concerning objective of "the optimal bandwidth division among the QoS classes", the decentralized theme applied in this part provided a more optimal division scheme among the QoS classes. This resulted from the fact that the division in this part was handseled by the NAs, where each NA was responsible for it outgoing links, and so, know better how to divide their resources according to their real-time needs.

Publications

Below is the list of publications delivered from the work related to this thesis:

Journals

Accepted and Published

- A. N. Quttoum, H. Otrok, Z. Dziong, *A collusion-resistant mechanism for autonomic resource management in Virtual Private Networks*, published by the Elsevier Journal of Computer Communications, Vol. 33, Issue 17, August 2010, Pages: 2070-2078.
- A. N. Quttoum, A. Jarray, H. Otrok, Z. Dziong, *An Optimal Dynamic Resource Partitioning Auction Model for Virtual Private Networks*, accepted for publication by the Springer Journal of Telecommunication Systems, September 2011.

Submitted

- A. Jarray, A. N. Quttoum, H. Otrok, Z. Dziong, *DDP: A Dynamic Dimensioning and Partitioning Model for Virtual Private Networks*, submitted to the Elsevier Press, Journal of Computer Communications, May 2011.
- A. N. Quttoum, A. Jarray, H. Otrok, Z. Dziong, *DDRP: A Decentralized Dynamic Resource Partitioning Model with Double-Auction for VPNs*, submitted to the Elsevier Press, Journal of Network and Computer Applications, August 2011.

Conferences

Accepted and Published

- A. N. Quttoum, H. Otrok, Z. Dziong, *ARMM: an Autonomic Resource Management Mechanism for virtual private networks*, in Proceedings of the Seventh Annual IEEE Consumer Communications and Networking Conference, CCNC, Las Vegas, Nevada, USA, January 9-12 2010.

Future Work

As a future work, the following can be considered:

- In the decentralized partitioning model, the work can be extended to address a pricing scheme. Through such pricing scheme, the cooperative NAs can be rewarded either by additional incentives, or provisioning price reduction.
- To resist collusion among the NAs in the decentralized partitioning model, i.e., all NAs may collude and offer high selling prices, the network controller may calculate a higher sell-threshold.

BIBLIOGRAPHY

Appleby, K. 2001. Oceano: Sla based management of a computing utility. *Proceedings of the IEEE/IFIP International Symposium on Integrated Network Management*, p. 855 – 868. IEEE.

Awduche, D., L. Berger, T. Li D. Gan, V. Srinivasan, et G. Swallow. 2001. “RSVP-TE: Extensions to RSVP for LSP Tunnels”. *The Internet Society*, p. 1–61.

Badia, L., M. Lindstrom, J. Zander, et M. Zorzi. 1997. Demand and pricing effects on the radio resource allocation of multimedia communication systems. *Proceeding of the IEEE GLOBECOM*, volume 7, p. 4116–4121. IEEE.

Barron, E. N., 2007. *Game Theory: An Introduction*. Wiley-Interscience.

Bellman, R. A., 1957. *Dynamic Programming*. New Jersey, NJ, USA : Princeton University Press.

Belzarena, P., A. Ferragut, et F. Paganini. 2009. Network bandwidth allocation via distributed auctions with time reservations. *Proceedings of the IEEE INFOCOM*, p. 1–5, Hawaii, 2009. IEEE.

Blake, S., D. Black, M. Carlson, E. Davies, Z. Wang, et W. Weiss. December 1998. “An architecture for differentiated services”. *Network Working Group RFC 2475*, 11(5).

Borst, S. C. et D.Mitra. June 1998. “Virtual Partitioning for Robust Resource Sharing: Computational Techniques for Heterogeneous Traffic”. *IEEE Journal on selected areas in communications*, p. 668–678.

Bouillte, E., D.Mitra, et K.G. Ramakrishnan. June 2002. “The structure and management of service level agreements in networks”. *IEEE Journal on selected areas in communications*, p. 691–699.

Cao, L. et H. Zheng. 2005. Distributied spectrum allocation via local bargaining. *in Proceedings of the Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*. IEEE SECON.

Carapinha, J. et J. Jimenez. 2009. Network virtualization: a view from the bottom. In Proceedings of the 1st ACM workshop on Virtualized Infrastructure Systems and Architectures, VISA ’09.

Cheng, Y., W. Zhuang, A. Leon-Garcia, et R. Q. Hu. Nov 2005. Efficient resource allocation for sla based wireless/wireline interworking. volume 1, p. 561–570. In Proceedings of the 2nd International Conference on Broadband Networks.

Cheng, Y., R. Farha, M. Sup Kim, A. Leon-Garcia, et J. Won-Ki Hong. Nov 2006a. “A generic architecture for autonomic service and network management”. *Computer Communications, Elsevier*, p. 3691–3709.

Cheng, Y., W. Song, et W. Zhuang. May 2006b. “Efficient Resource Allocation for Policy-Based Wireless/Wireline Interworking”. *Mobile Networks and Applications*, p. 661–679.

Chinneck, J. W., 2001. *Practical Optimization: a Gentle Introduction*. Carleton, ON, Canada : Carleton University.

Chou, C. T. 2004. “Traffic engineering for MPLS-based virtual private networks”. *Elsevier, Computer Networks*, p. 319–333.

Chowdhury, N. M. M. K. et R. Boutaba. 2010. “A survey of network virtualization”. *Computer Networks*, p. 862–876.

Clarke, E. H. 1971. “Multipart pricing of public goods”. *Springer, Public Choice*, p. 17–33.

Cohen, R. et G. Kaempfer. Dec 2000. “On the Cost of Virtual Private Networks”. *IEEE/ACM Transactions on Networks*, p. 775–784.

Dantzig, G. B., 1963. *Linear programming and extensions, Princeton Landmarks in Mathematics*. New Jersey, NJ, USA : Princeton University Press.

Davie, B. et Y. Rekhter, 2000. *MPLS: technology and applications*. San Francisco, CA, USA : Morgan Kaufmann Inc.

Dramitinos, M., George D. Stamoulis, et C. Courcoubetis. 2007. “An auction mechanism for allocating the bandwidth of networks to their users”. *Computer Networks*, p. 4979–4996.

Duffield, N. G., P. Goyal, A. Greenberg, P. Mishra, K. Ramakrishnan, et J. van der Merive. 1999a. A flexible model for resource management in virtual private networks. *Proceedings of the SIGCOMM ’99*, p. 95–108. ACM.

Duffield, N. G., P. Goyal, A. Greenberg, P. Mishra, K. Ramakrishnan, et J. van der Merwe. 1999b. A flexible model for resource management in virtual private networks. *in Proceedings of the of the conference on Applications, technologies, architectures, and protocols for computer communication*, p. 95–108. IEEE, SIGCOMM.

Duffield, N. G., P. Goyal, A. Greenberg, P. Mishra, K. Ramakrishnan, et J. van der Merwe. 2002. “Resource management with hoses: point-to-cloud services for virtual private networks”. *IEEE/ACM Transactions on Networking*, p. 679–692.

Dziong, Z., 1997. *ATM Network Resource Management*. New York : McGraw-Hill.

Dziong, Z., M. Pioro, U. Korner, et T. Wickberg. 1988. On adaptive call routing strategies in circuit switched networks maximum revenue approach. In Proceedings of the 11th International Teletranc Congress, Torino, Italy.

El-Kadi, M., S. Olariu, et H Abdel-Wahab. Nov 2002. “ A Rate-Based Borrowing Scheme for QoS Provisioning in Multimedia Wireless Networks ”. *IEEE Transaction on parallel and Distributed Systems*, p. 3691–3709.

Epstein, D. 1994. Finding the k shortest paths. p. 154–165. In Proceedings of the 35th Annual Symposium on Foundations of Computer Science.

Erlebach, T. et M. Ruegg. 2004. Optimal bandwidth reservation in hose model vpns with multi-path routing. *in Proceedings of the twenty-third Annualv Joint Conference of the IEEE Computer and Communications Societies*, volume 1, p. 2275–2282. IEEE, INFOCOM.

Etkin, R., A. Parekh, et D. Tse. 2007. “ Spectrum sharing for unlicensed bands ”. *IEEE Journal of Selected Areaa in Communcation*, p. 517–528.

Farha, R. et A. Leon-Garcia. 2006. Blueprint for an autonomic service architecture. *Proceedings of the* 2^{nd} *ICAS*.

Firestone, S., T. Ramalingam, et S. Fry. 2007. “ Voice and Video Conferencing Fundamentals ”. *Cisco Press*, p. 1–480.

Gandhi, S., C. Buragohain, L. Cao, H. Zheng, et S. Suri. 2007. A general framework for wireless spectrum auctions. *in Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, p. 22–33, Dublin, 2007. IEEE DySPAN.

Ghobadi, M., S. Ganti, et G. C. Shoja. 2007. Hierarchical provisioning algorithm for virtual private networks using the hose model. *in Proceedings of the Global Telecommunications Conference*, p. 2467–2471. IEEE GLOBECOM.

Ghobadi, M., S. Ganti, et G. C. Shoja. 2008. “ Resource optimization algorithms for virtual private networks using the hose model ”. *ELSEVIER, Computer Networks*, p. 3130–3147.

Graupner, S., A. Andrzejak, V. Kotov, et H. Trinks. 2003. Adaptive control overlay for service management. *Proceedings of the Workshop on the Design of Self-Managing Systems*. IEEE.

Groves, T. 1973. “ Incentives in Teams ”. *Econometrica*, p. 617–631.

Gupta, A., A. Kumar, et T. Roughgarden. 2003. Simpler and better algorithms for network design. *in Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. IEEE, STOC ’03.

Han, Z., C. Pandana, et K. J. R. Liu. 2007. Distributive opportunistic spectrum access for cognitive radio using correlated equilibrium and no-regret learning. *in Proceedings of the IEEE Wireless Communications and Networking Conference*, p. 11–15, Hong Kong, 2007. IEEE WCNC.

Haung, Y. et J. Ho. Nov 2002. “ Distributed call admission control for a heterogeneous PCS network ”. *IEEE Transaction on Computing*, p. 1400–1409.

Howard, R. A., 1960. *Dynamic Programming and Markov Process*. Cambridge, Massachusetts, USA : The M.I.T Press.

HP-Laboratories. Jul 2003. The hp vision for the adaptive enterprise: achieving business agility. HP.

Huang, J., R. Berry, et M. L. Honig. 2006. “ Auction-based spectrum sharing ”. *ACM Mobile Networks and Applications MONET*, p. 405–408.

Ileri, O., D. Samardzija, et N. B. Mandayam. 2005. Demand responsive pricing and competitive spectrum allocation via a spectrum server. *in Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, p. 194–202, Baltimore, MD, USA, 2005. IEEE DySPAN.

ITUT-Rec-P800. 2000. The e-model, a computational model for use in transmission planning. GE.

Jeffrey, K., MacKie-Mason, et Hal R. Varian. 1994. Pricing the internet. Computational Economics, 9401002, EconWPA.

Ji, Z. et K. J. R. Liu. 2006. Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users. *in Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, Reston, VA, USA, 2006. IEEE.

Ji, Z. et K. J. R. Liu. 2007. “ Dynamic Spectrum Sharing: A Game Theoretical Overview ”. *IEEE Communication Magazine*, p. 88–94.

Ji, Z. et K. J. R. Liu. 2008. “ Multi-stage pricing game for collusion-resistant dynamic spectrum allocation ”. *IEEE Journal of Selected Areas in Communications*, p. 182–191.

Juttner, A., I. Szabo, et A. Szentesi. 2003. On bandwidth efficiency of the hose resource management model in virtual private networks. *in Proceedings of the twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, volume 1, p. 386–395. IEEE, INFOCOM.

Kelly, F. 1997a. “ Charging and rate control for elastic traffic ”. *European Transactions on Telecommunications*, p. 33–37.

Kelly, F. P. 1988. “ Routing in Circuit-Switched Networks: Optimization, Shadow Prices and Decentralization ”. *J stor*, p. 112–144.

Kelly, F. P. 1997b. Charging and accounting for bursty connections. p. 253–278. Internet Economics, MIT Press Cambridge, MA, USA.

Kim, M. S., A. Tizghadam, A. Leon-Garcia, et J.W.-K. Hong. 2005. Virtual network based autonomic network resource control and management system. *in Proceedings of the Global Telecommunications Conference*. IEEE GLOBCOM.

Kumar, A., R. Rastogi, A. Silberschatz, et B. Yener. 2001. Algorithms for provisioning virtual private networks in the hose model. *in Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, San Diego, California, USA, 2001. IEEE SIGCOMM 01.

Kumar, A., R. Rastogi, A. Silberschatz, et B. Yener. 2002. “ Algorithms for provisioning virtual private networks in the hose model ”. *IEEE/ACM Transaction on Networking*, p. 565–578.

Lai, F. S., J. Misic, et S. T. Chanson. 1998. Complete sharing versus partitioning: quality of service management for wireless multimedia networks. In Proceedings of the 7th International Conference on Computer Communications and Networks.

Lazar, A. et N. Semret. 1999. Design and analysis of the progressive second price auction for network bandwidth sharing. Telecommunication Systems, Special Issue in Network Economics.

Malla, A., M. El-Kadi, S. Olariu, et P. Todorova. Nov 2003. “ A Fair Resource Allocation Protocol for Multimedia Wireless Networks ”. *IEEE Transaction on parallel and Distributed Systems*, p. 3691–3709.

Mark, J. W., T. C. Wong, M. Jin, B. Bensaou, et K. C. Chua. 2000. Resource allocation for multiclass services in cellular systems. In Proceedings of the IEEE International Conference in Communication Systems.

Mas-Colell, A., M. Whinston, et J. Green, 1995. *Microeconomic Theory*. New York, NA, USA : Oxford University Press.

Microsoft-Corporation. 2004. Dynamic systems initiative overview. *White Paper*. Microsoft.

Mohamed, N., H. Otrok, L. Wang, M. Debbabi, et P. Bhattacharya. 2011. “ Mechanism Design-Based Secure Leader Election Model for Intrusion Detection in MANET ”. *IEEE TDSC*, p. 89–103.

Morris, P., 1994. *Introduction to Game Theory*. Springer, 1st edition.

Myerson, R., 1991. *Game Theory: Analysis of Conflict*. USA : Harvard University Press.

Nash, J. 1951. “ Non-cooperative games ”. *Annals of Mathematics, Second Series*, p. 286–295.

Niebert, N., I. El Khayat, S. Baucke, R. Keller, R. Rembarz, et J. Sachs. 2008. “ Network Virtualization: A Viable Path Towards the Future Internet ”. *Wireless Press Communications*, p. 511–520.

Nisan, N. et A. Ronen. 2007. “Computationally Feasible VCG Mechanisms”. *Journal of Artificial Intelligence Research*, p. 19–47.

Nisan, N. et Amir Ronen. 1999. Algorithmic mechanism design (extended abstract). *Proceedings of the* 31 *STOC*. ACM.

Niyato, D. et E. Hossain. 2007. “Integration ofWiMAX andWiFi: optimal pricing for bandwidth sharing”. *IEEE Communication Magazine*, p. 140–146.

Puterman, M. L., 1994. *Markov Decision Processes*. New York, NY, USA : John Wiley and Sons.

Quttoum, A., H. Otrok, et Z. Dziong. 2010a. ARMM: An autonomic resource management mechanism for virtual private networks. *in Proceedings of the 7th IEEE Consumer Communications and Networking Conference, CCNC*, p. 1–6, Las Vegas, Nevada, USA, 2010a. IEEE.

Quttoum, A. N., H. Otrok, et Z. Dziong. Sep 2010b. “A Collusion Resistant Mechanism for Autonomic Resource Management in Virtual Private Networks.”. *In the Special Issue of the Elsevier Computer Communications Journal on Applied Sciences in Communication Technologies*, p. 2070–2078.

Rixner, S. 2008. “Network Virtualization: Breaking the Performance Barrier”. *Queue Virtualization Magazine*, p. 37–52.

Semret, N., R. Liao, A. Campbell, et A. Lazar. 2005. “Pricing, provisioning and peering: Dynamic markets for differentiated internet services and implications for network interconnections”. *IEEE JSAC*, p. 78–118.

Stenmark, D. 2002. Information vs. knowledge: the role of intranets in knowledge management. p. 928–937. In Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE, HICSS.

Sup, K. M., T. Ali, A. Leon-Garcia, et H. James Won-Ki. 2005. Virtual network based autonomic network resource control and management system. *Proceedings of the IEEE GlobCom*. IEEE.

Thompson, K., G. J. Miller, et R. Wilder. Nov 1997. “Wide-area Internet traffic patterns and characteristics”. *IEEE Network*, p. 10–23.

Tran, C. et Z. Dziong. 2010. “Service overlay network capacity adaptation for profit maximization”. *IEEE Transactions on Network and Service Management*, p. 72–82.

Tran, C., Z. Dziong, et M. Pioro. 2007. Service overlay network capacity adaptation for profit maximization. *Networking 2007. 6th International IFIP-TC6 Networking Conference Proceedings*, volume 4479, p. 691–702. Springer Heidelberg.

Vickrey, W. 1961. “Counterspeculation, auctions, and competitive sealed tenders”. *Journal of Finance*, p. 8–37.

Wang, B., Z. Ji, et K. J. R. Liu. 2007. Self-learning repeated game framework for distributed primary-prioritized dynamic spectrum access. *in Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, p. 631–638. IEEE SECON.

Wang, B., Y. Wu, Z. Ji, K. J. Ray Liu, et T. Clancy. 2008. “Game Theoretical Mechanism Design for Cognitive Radio Networks with Selfish Users”. *IEEE Signal Processing Magazine*, p. 74–85.

Williams, A. G. 1965. “On Stochastic Linear Programming”. *Journal of the Society for Industrial and Applied Mathematics*, 13(4).

Yao, J., J. M. Mark, T. C. Wong, Y. H. Chew, K. M. Lye, et K. C. Chua. Nov 2004. “Virtual partitioning resource allocation for multiclass traffic in cullular systems with QoS constraints”. *IEEE Transaction on Vehicular Technology*, p. 847–864.