DYNAMIC WAVELENGTH ALLOCATION IN IP/WDM METRO ACCESS NETWORKS

A DISSERTATION SUBMITTED TO
THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Emre Yetginer
June, 2008

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Assoc. Prof. Dr. Ezhan Karaşan(Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Prof. Dr. Erdal Arıkan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Prof. Dr. Semih Bilgen

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Assoc. Prof. Dr. Nail Akar

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of doctor of philosophy.

Assoc. Prof. Dr. Mustafa Akgül

Approved for the Institute of Engineering and Science:

Prof. Dr. Mehmet B. Baray
Director of the Institute

ABSTRACT

DYNAMIC WAVELENGTH ALLOCATION IN IP/WDM METRO ACCESS NETWORKS

Emre Yetginer
Ph.D. in Electrical and Electronics Engineering
Supervisor: Assoc. Prof. Dr. Ezhan Karaşan
June, 2008

Increasing demand for bandwidth and proliferation of packet based traffic have been causing architectural changes in the communications infrastructure. In this evolution, metro networks face both the capacity and dynamic adaptability constraints. The increase in the access and backbone speeds result in high bandwidth requirements, whereas the popularity of wireless access and limited number of customers in metro area necessitates traffic adaptability. Traditional architecture which has been optimized for carrying circuit-switched connections, is far from meeting these requirements. Recently, several architectures have been proposed for future metro access networks. Nearly all of these solutions support dynamic allocation of bandwidth to follow fluctuations in the traffic demand. However, reconfiguration policies that can be used in this process have not been fully explored yet. In this thesis, dynamic wavelength allocation (DWA) policies for IP/WDM metro access networks with reconfiguration delays are considered. Reconfiguration actions incur a cost since a portion of the capacity becomes idle in the reconfiguration period due to the signalling latencies and tuning times of optical transceivers. Exact formulation of the DWA problem is developed as a Markov Decision Process (MDP) and a new cost function is proposed to attain both throughput efficiency and fairness. For larger problems, a heuristic approach based on first passage probabilities is developed. The performance of the method is evaluated under both stationary and non-stationary traffic conditions. The effects of relevant network and traffic parameters, such as delay and flow size are also discussed. Finally, performance bounds for the DWA methods are derived.

Keywords: Metro Access Networks, IP over WDM, Dynamic Wavelength Allocation, Reconfiguration, Markov Decision Process, Reconfiguration Delay.

ÖZET

IP/WDM METRO ERİŞİM AĞLARINDA DİNAMİK DALGABOYU TAHSİSİ

Emre Yetginer
Elektrik ve Elektronik Mühendisliği, Doktora
Tez Yöneticisi: Doç. Dr. Ezhan Karaşan
Haziran, 2008

Bant genişliği talebindeki artış ve paket tabanlı trafiğin çoğalması iletişim altyapısında mimari değişikliklere sebep olmaktadır. Bu evrimde metro ağları hem kapasite hem de dinamik uyarlanabilirlik kısıtları ile karşı karşıya bulunmaktadır. Erişim ve omurga hızlarındaki artış yüksek bant genişiliği ihtiyacını doğururken, kablosuz ağların popülerliği ve metro alanındakı kısıtlı kullanıcı sayısı trafiğe uyarlanabilirliği gerektirmektedir. Çevrim anahtarlamalı bağlantılar için eniyilenmiş olan geleneksel mimari bu ihtiyaçları karşılamaktan oldukça uzaktır. Son zamanlarda, yeni nesil metro erişim ağları için çeşitli mimariler önerilmiştir. Bu çözümlerin hemen tamamı trafik talebindeki dalgalanmaları takip etmek için bant genişiliğinin dinamik olarak tahsisini desteklemektedir. Ancak, bu süreçte kullanılabilecek yeniden düzenleşim politikaları henüz tam anlamıyla araştırılmamıştır. Bu tezde, düzenleşim gecikmesine sahip IP/WDM ağları için dinamik dalgaboyu tahsisi (DWA) politikaları ele alınmaktadır. Sinyalleşme gecikmeleri ve optik verici-alıcıların akortlanma zamanlarından dolayı düzenleşim süresi boyunca kapasitenin bir bölümünün atıl kalması, yeniden düzenleşim eylemleri için bir maliyet oluşturmaktadır. DWA probleminin kesin formülasyonu bir Markov Karar Süreci (MDP) olarak geliştirilmiş ve hem debi etkinliğini hem de servis adilliğini sağlayacak yeni bir maliyet fonksiyonu önerilmiştir. Büyük problemler için, ilk geçiş olasılıkları tabanlı buluşsal bir yöntem geliştirilmiştir. Yöntemin başarımı hem durağan hem de durağan olmayan trafik koşullarında değerlendirilmiştir. Gecikme ve akış boyutu gibi ilgili parametrelerin etkileri de tartışılmıştır. Son olarak, DWA metotları için performans sınırları derlenmiştir.

Anahtar sözcükler: Metro Erişim Ağları, IP/WDM, Dinamik Dalgaboyu Tahsisi, Yeniden Düzenleşim, Markov Karar Süreçleri, Düzenleşim Gecikmesi.

To my beloved wife,
Esin

Acknowledgement

I would like to express my sincere gratitude and appreciation to my supervisor, Assoc. Prof. Dr. Ezhan Karaşan, for his invaluable mentorship, suggestions, and encouragement throughout the development of this thesis. I am greatly indebted to him for his confidence in me and personal guidance in all stages of my graduate education.

Special thanks to Prof. Dr. Erdal Arıkan, Prof. Dr. Semih Bilgen, Assoc. Prof. Dr. Nail Akar and Assoc. Prof. Dr. Mustafa Akgül for reading and commenting on the thesis.

I would also like to thank TUBITAK-UEKAE, and Assoc. Prof. Dr. S. Gökhun Tanyer and Mr. Önder Yetiş for their support and encouragement during my graduate studies.

Finally, I would like to thank my family for their life-long love and constant support. I would like to give special thanks to my wife Esin whose patient love enabled me to complete this work.

Contents

List of Figures

List of Tables

List of Abbreviations

AN	Access Node
ATM	Asynchronous Transfer Mode
CO	Central Office
DSL	Digital Subscriber Line
DWA	Dynamic Wavelength Allocation
DWDM	Dense Wavelength Division Multiplexing
EPON	Ethernet PON
FS	Flow Sum
ILP	Integer Linear Programming
IP	Internet Protocol
LAN	Local Area Network
MAC	Medium Access Protocol
MDP	Markov Decision Process
MILP	Mixed Integer Linear Programming
NFS	Normalized Flow Sum
NGS	Next Generation SONET

NSFS	Normalized Squared Flow Sum
PON	Passive Optical Network
RPR	Resilient Packet Ring
RTT	Round-Trip Time
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical Network
SWA	Static Wavelength Allocation
TCP	Transport Control Protocol
TDM	Time Division Multiplexing
WDM	Wavelength Division Multiplexing

Chapter 1

Introduction

Communication networks are composed of three major layers: access, metro and long-haul. Access (distribution) networks provide the last mile connectivity for residential and business users. Several different technologies have been used to establish the link between the customers and the service provider including dial-up, DSL (Digital Subscriber Line), cable TV, wireless LAN and PON (Passive Optical Networks). The access network is terminated at a central office (CO) owned by the service provider. In general, a CO serves a district of a town and the COs in a town are connected to each other to form the metro access network. A specific CO at each metro access network is designated as the hub CO. Metro core network connects hub COs to each other and may cover the whole city. The connections between metro networks are through the long-haul (core) network consisting of intercity and regional links.

The steady increase of the Internet traffic has caused architectural and conceptual changes in communication networks. The infrastructure, once designed to carry legacy voice services, is no longer able to put up with this ever increasing packet-based traffic. Long-haul backbone networks have been adapted to this change using the optical transmission technology and dense wavelength division multiplexing (DWDM), which enables concurrent transmission of more than 100 channels each at 10 Gbps over a single fiber. In the future, core networks are expected to evolve towards a fully optical transport network architecture [1].

Meanwhile, in the access side service rates have increased to levels in the order of 10 Mbps. With the penetration of optical fibers down to the premises of end users, the target is to offer gigabit per second rates directly to the customers [2]. But, metro networks, that are in between access and core networks lag behind in terms of speed and capacity. Hence, they constitute a barrier for this large volume of traffic to be transmitted from access networks to the high speed backbone networks. The pressure from the access side forces metro networks, most of which still rely on legacy time division multiplexing (TDM) based technologies, into an evolutionary process [3]. High capacity, protocol transparency, cost efficiency and dynamic traffic adaptability are major issues in this transformation [4]. Most of the solutions designed for future metro access networks (e.g., Next Generation SONET (NGS) [5], IP/WDM [6]) support dynamic reconfiguration in order to meet cost efficiency and traffic adaptability requirements. Likewise, reconfigurability is also possible for Ethernet Passive Optical Networks (EPON) that are seen as a promising technology for future access networks [7]. However, development of the methods that can be used for dynamic reconfiguration is still an open research problem.

In this thesis, a wavelength routed IP/WDM ring network ([8, 9, 10, 11]) is considered. This architecture is most suitable for metro access networks with hubbed traffic patterns, where the local traffic between access nodes is negligible. Hence, a centralized control is implemented at the hub node. On the other hand, for metro networks with more homogeneous traffic patterns, a packet based IP/WDM ring ([12, 13, 14]) may be the preferred solution. The main focus of this work is to discover the trade-offs and potential benefits of dynamic capacity allocation, and develop reconfiguration policies for efficient capacity utilization.

1.1 IP/WDM Network Architecture and the DWA Problem

An IP/WDM metro access network is constructed by connecting access nodes (each located at a CO) and the hub node (located at the hub CO) in a ring

topology. The ring may consist of a single fiber or multiple fibers where a fiber is capable of supporting tens of wavelengths. The hub node forms lightpaths to each access node (AN) by allocating separate wavelengths. Therefore, the logical network has a tree topology. Traffic from the distribution networks are aggregated at the corresponding ANs and forwarded to the hub node using the established lightpaths. Each AN may also be equipped with tunable receivers and transmitters in which case the wavelength allocation and hence the logical topology can be dynamically changed. As a result, two different approaches may be considered for wavelength allocation in these networks: static wavelength allocation and dynamic wavelength allocation.

In *static wavelength allocation*, traffic demand is measured and/or predicted at each node of the network and available wavelengths are allocated in accordance with the traffic projections. The resulting capacity allocation is not changed in time. This approach is commonly used in core networks where large volumes of traffic aggregation results in slowly changing and hence to a large extent stable and predictable traffic patterns [15]. Therefore, static design of the logical topology and over-provisioning the capacity to handle traffic uncertainty prove to be sufficient. Reconfiguration is mostly manual and requires a time duration in the order of hours or days but this is not a concern since reconfiguration is required only in case of large and persistent demand deviations, such as the addition of new nodes to the network or network failures. However, the proximity of metro networks to the end users differentiates them from core networks. The limited number of users served results in low traffic aggregation levels and frequent fluctuations in the traffic demand. Increasing bandwidth and popularity of wireless access solutions further contribute to the traffic uncertainty and variability. Since each node of a metro access network serves a different district of a town, it is possible to observe nearly periodic oscillations in the traffic demand [16], [17]. These variations may occur at different time scales. Traffic patterns may change on a daily basis, e.g., in weekdays and weekends different portions of the network may become congested. During working hours, hot spots may shift from residential areas to business districts, corresponding to a traffic variation on the order of hours. At the extreme case, where traffic aggregation is very low, individual flows

corresponding to high-speed transactions may cause more frequent fluctuations. As a result, peak traffic demand at each access node may be much larger than the average rate. A static wavelength allocation that is made based on average traffic forecasts may cause congestion at times of peak demand. On the other hand, an allocation strategy based on peak traffic rates at each node may be a waste of bandwidth most of the time.

In the *dynamic wavelength allocation (DWA)* scheme, traffic at each node is monitored and the wavelength allocation is changed in accordance with the demand variations. This approach may result in significant improvements in efficiency and fairness with respect to the static allocation when the demand deviations are large and frequent, as in metro access networks. The idea may be demonstrated on a hypothetical example. Consider a network consisting of just 2 ANs and let B denote the capacity of a single wavelength channel and D_i be the offered traffic at node i. Suppose that the $D_1 = 2B$ and $D_2 = 6B$ in working hours and $D_1 = 6B$ and $D_2 = 2B$ in the rest of the day. To satisfy a target utilization level of 66%, with static wavelength allocation each node has to be allocated 9 wavelength channels and a total of 18 wavelengths are required. However, with dynamic wavelength allocation 12 wavelengths are sufficient to achieve the desired utilization level.

Besides these potential benefits, reconfiguration of wavelengths has an associated cost. Due to the signalling requirements and latencies related to the tuning of transmitters and receivers, the wavelengths that are being reconfigured become unavailable for a certain duration. The presence of this delay introduces a trade-off between the reconfiguration costs and the responsiveness of the network to the traffic changes. Hence, switching of a wavelength should be performed only if the expected long-term benefits overweighs the cost of reconfiguration. That is, a poor switching decision may require another reconfiguration action at the very next decision point, incurring another cost. Therefore, switching decisions should consider not only the immediate benefit that will be obtained during the next time interval but also the long-term effects on the future reconfigurations and demand-capacity match.

Wavelength allocation in IP/WDM access networks belongs to the general class of resource allocation problems. In such problems, there are a set of limited resources and a set of demands each requiring a specified amount of resources. Utilization of a resource has an associated cost and completion of demands result in revenue. Alternatively, each unsatisfied demand in the system may introduce a cost. The goal of resource allocation is to minimize the total cost or maximize the revenue with the distribution of resources among competing alternatives. There are numerous instances of the resource allocation problem in diverse fields. Sharing of a society's resources among its members, distribution of time slots to tasks in a computer system, assigning resources such as CPUs and memory to different jobs in a computing facility (e.g., a service grid, a data center or a multi-processor machine), bandwidth allocation in computer networks, production planning and portfolio selection are just a few examples where a set of resources is required to be distributed among a set of entities or activities.

Basically, resource allocation problems can be divided into two categories based on the time and cost associated with the migration of resources. In the first and simpler case, the resources can be migrated instantaneously and with no cost. Then, it would be sufficient to reassign resources in response to changing conditions to maximize revenue or minimize the cost [18]. However, if the observation of system conditions are infrequent, then it is required to use a forward-looking reassignment policy which evaluates the expected gains and losses during the next time interval. The second category of resource allocation problems considers the more realistic case where resource migrations require a non-negligible time during which resources are idling (or are not fully utilized) or results in some cost. For this case, resource migration decisions should consider not only the immediate benefit that will be obtained during the next time interval but also the long-term effects. Unfortunately, even for the simplest problems of this type, it is hard to characterize the optimum policy explicitly [19, 20]. The DWA problem considered in this thesis belongs to the latter class discussed above due to the presence of reconfiguration delays associated with wavelength switching actions.

1.2 Main Contributions

Despite its importance, dynamic resource allocation strategies in the context of metro access networks have not been comprehensively studied in the literature. Wavelength routed IP/WDM networks are considered in several work. In [21, 22, 23] multi-hop ring architecture is considered, and in [24] a simple heuristic method is proposed for single hop networks. For packet mode IP/WDM networks a heuristic method based on Markov Decision Process (MDP) formulation is developed in [25] under some restrictive assumptions on the traffic process.

The first contribution of this thesis is the exact formulation of the DWA problem as an MDP, the solution of which results in the optimum switching policy. A new cost function is also proposed, which jointly achieves slowdown and fairness objectives. The cost function has some useful properties and well suited for problems where load balancing between servers is desired. The superiority of this function to alternative cost definitions found in the literature is also demonstrated.

Another contribution of this thesis is the development of a new heuristic method for the DWA problem. It aims to minimize the cost function mentioned above. The novelty of this heuristic lies in the usage of first passage probabilities to assign quantitative values to possible reconfiguration actions. With this approach, the capacity wasted during the reconfiguration period, hence the reconfiguration cost, is implicitly taken into account in a natural way. The heuristic method also provides a hysteresis region automatically. As a result, the heuristic performs well under a wide range of network and traffic parameters without any modifications. Finally, an efficient method for the implementation of the heuristic is developed which relies on the off-line calculation of some representative values. With this approach, the heuristic method reduces to a set of simple table look-up and comparison operations suitable for real-time usage. The approximation error is bounded analytically and can be made arbitrarily small by increasing the size of the look-up tables.

In this work, flows correspond to large scale transactions or aggregation of

small flows, not to individual short connections. Therefore, the average size of the flows is large and inter-arrival times are chosen correspondingly. The solution is based on the assumption that the bottleneck is at the metro access network. This is a reasonable assumption when the high capacity of the backbone networks is considered. The possibility of having the bottleneck at the distribution network is also small, since users with very large traffic demands are directly connected to access nodes. Without this assumption, the DWA turns into a much more complicated problem for which a solution is hard to obtain.

For simplicity, the flows are assumed to be elastic, which may not ideally hold for TCP flows. The implications of this fact for the real life flows are analyzed and the modification of the heuristic method to handle this behavior is identified as a further research area.

In order to be able to use the MDP approach, the traffic process is assumed to be Markovian, and Poisson flow arrivals and exponential flow sizes are used for simplicity. This is indeed not a realistic assumption but makes the interpretation of the results easier. The current work can be extended to more realistic models of traffic by using more complicated Markovian models for the arrival process, such as Markov Modulated Poisson Process (MMPP). It is also possible to use general distributions for flow sizes (e.g., Pareto distribution) and formulating the problem as a semi-Markov process.

1.3 Overview of the Thesis

To formulate the DWA problem, a basic framework for IP/WDM networks is developed in this thesis. A multi-point to point traffic pattern is assumed and switching actions are performed at flow arrival and departure instants. At most one wavelength is allowed to be in the switching state. Due to connectivity requirements, it is enforced to have at least one wavelength at each node at any time. It is also assumed that the flows are elastic and each node is equipped with a packet scheduler so that a processor sharing model is applicable for each AN.

Using the framework developed and with appropriate definitions of state representation, action space and transition rates, the DWA problem is formulated as a continuous-time MDP. The optimum DWA policy is obtained using dynamical programming methods to solve the MDP. The performance of the resulting policy directly depends on the cost function used in the MDP formulation. A novel cost function (NSFS) which takes into account both the throughput and fairness objectives is developed. Two other cost functions (FS and NFS) derived from the literature are also utilized to obtain three optimum DWA policies. These policies are compared on a 3-node test network under stationary traffic, and it is demonstrated that the proposed cost function has a superior performance in terms of both slowdown and fairness. The results for the static wavelength allocation policy are also obtained and used for comparison purposes. It is observed that all the DWA policies improve the slowdown and fairness. Among the dynamic policies, best performance is achieved by the NSFS at all network load levels. The slowdown is improved by 25% to 35% with respect to static allocation. With increasing network load, relative performance of NFS decreases while FS and NSFS achieves larger improvements. It is also shown that NSFS attains an impressive improvement in fairness.

Since the MDP solution of the DWA problem is feasible only for small networks, an efficient heuristic approach (HM3), based on the cost function NSFS and utilizing first passage probabilities, is proposed. The method inherently takes into account the delays associated with the reconfiguration actions. It is shown that HM3 performs close to the optimum policy in terms of slowdown. The optimality gap is below 5% for moderate load and it decreases further as the network load increases. Two other heuristic policies proposed in the literature (HM1 and HM2) are also adapted to the problem. It is observed that HM3 achieves superior performance with respect to HM1 and HM2, for the whole range of network load. The heuristic methods are then compared using a 5-node test network under non-stationary traffic. The results suggest that HM3 achieves the best performance by realizing minimum number reconfigurations. The slowdown is nearly halved with respect to the static policy. The improvement is 35% and 9% compared to HM1 and HM2, respectively. HM3 is also shown to achieve maximum fairness.

The effects of several traffic and network parameters on the performance of the methods are also analyzed. First, average flow size and channel bandwidth are discussed. The results clearly indicate that only HM3 is successful at the short flow length regime. HM1 and HM2 perform even worse than the static policy as the average flow size gets smaller. Secondly, average reconfiguration delay is considered, and it is shown that the performance of HM1 and HM2 decrease below the static policy as the reconfiguration delay takes larger values. HM3 is able to select appropriate actions and improve the performance for all values of reconfiguration delay. Finally, the behavior of the methods for networks with different number of wavelengths are analyzed, and the superiority of HM3 is observed for all cases studied.

Theoretical bounds on the performance improvement that can be achieved by DWA policies are also investigated. By relaxing some constraints of the DWA problem and approximating the system as a single M/M/1-PS queue, two lower bounds are obtained. The first bound, LB1, is obtained without considering the connectivity constraint and therefore assuming homogeneous departure rates. A tighter bound, LB2, is obtained by incorporating this constraint and constructing an inhomogeneous Markov chain. These bounds are demonstrated on a 3-node network, and it is observed that LB2 takes values which are 32-48% higher than LB1. Comparisons with the results of the optimum policy show that the bound is indeed 20-30% lower than the minimum results achievable.

The rest of the thesis is organized as follows. Chapter 2 provides brief information on the general network architecture, along with the developments and trends at each layer of the network. Chapter 3 introduces a more detailed view of the IP/WDM network and the framework used for the DWA problem. In Chapter 4, exact solution of the DWA problem is obtained through an MDP formulation. Three cost functions are considered and compared through numerical results. The heuristic reconfiguration policies are discussed in Chapter 5 and performance of these methods are compared in Chapter 6. Theoretical bounds on the DWA performance are developed in Chapter 7. Future research topics are identified in Chapter 8, and Chapter 9 concludes the thesis.

Chapter 2

Emerging Optical Transport Technologies in Access, Metro and Core Networks

Today's data networks are categorized mainly based on their geographic extents. In this classification, the communication infrastructure has the hierarchical layers of long-haul (backbone), regional (metro core), metropolitan (metro access), and access (distribution) networks as depicted in Figure 2.1. End-users are connected to access networks as shown at the bottom of the figure. These access networks are terminated at a Central Office (CO), which is a part of the metro access ring. A hub CO connects the metro access network to the regional (metro core) network. Finally, regional networks are all connected to the long haul core.

In the following sections, a basic overview of these network partitions is provided along with discussions on the developments and trends in each layer. Metro networks are considered in more detail and several proposals for future metro access networks are presented along with the WDM based solutions.

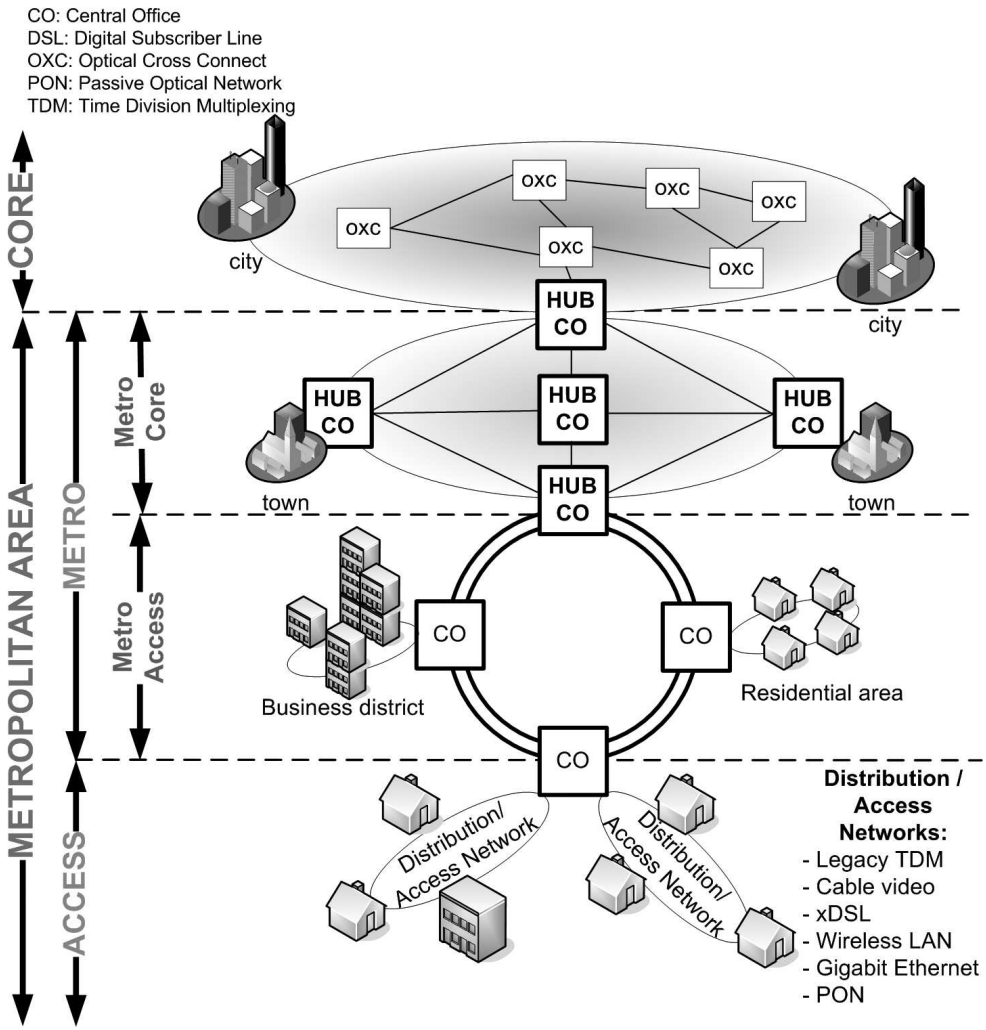


Figure 2.1: Communication network architecture.

2.1 Access Networks

Access networks cover a limited area (up to about 20 km) and provide "last mile" connectivity to business and residential customers. Therefore, they are also called as the last-mile networks. In recent years, the importance of this segment has been increased and it is renamed by the Ethernet community as the "first mile" to emphasize this importance.

The historical development of this area had been based on the circuit switched

services, such as telephony and cable TV. Hence, the conventional infrastructure is optimized for voice-like circuit oriented applications and it uses twisted pair and coaxial cable networks as physical medium. Dial-up services has been used for data networking.

In the Internet era, the demand has shifted to packet based data communications. Service providers react to this change by producing solutions using existing infrastructure. Telephone operators offered Digital Subscriber Line (DSL) services while TV operators came up with cable modems. The rates of these services have increased over time but higher rates are achieved at the cost of a shorter distribution range. Today, DSL is widely deployed and offers multi megabit speeds over copper.

Meanwhile, technological advances in wireless communications resulted in the introduction of wireless access options such as 3/3.5G wireless systems and IEEE 802.11 wireless LAN, which are also maturing to serve over megabit speeds. Wireless networking is still an area of research and technological development. Especially, free space optics and ad-hoc networking are popular subjects for both academic community and industry.

Increase of the access bandwidth results in new applications that demand more capacity. On the residential side, content rich applications and real time services coupled with the growth of Internet trigger the demand for higher bandwidth. Operators' goal of providing triple play services (bundled service package of voice, video and data) to their customers requires even higher bandwidths. On the business side, virtual private networking (VPN) services and storage area networking (SAN) applications increase the demand for data traffic. The overall result is the dominance of data (mostly IP/Ethernet based) traffic over voice and leased line services.

To increase the access bandwidths even further to Gbps level, current trend is to bring the optical fiber closer to the users. Indeed, this idea has existed for a couple of decades, but the the enabling factors, such as the availability of affordable optical components, deployment of fiber cables, readiness of service providers and development of sufficient bandwidth demand were missing. With

the realization of these factors, the concept of optical access is becoming a feasible option. With the penetration of the optical fiber to the access domain, it is expected to enable the delivery of any current and foreseeable set of broadband services.

Several alternatives exist for the optical access depending on the reach of the fiber: Fiber to the curb/cabinet (FTTC), fiber to the building (FTTB), and fiber to the home (FTTH) [26]. These are also collectively termed as FTTX. In all of these architectures, there is an Optical Line Termination (OLT) device at the central office, which serves as the access node to the metro network. And there is a corresponding Optical Network Unit (ONU) which is connected to the OLT over fiber cables. In FTTC solution, ONU is located at a curb/cabinet, and users are connected to this ONU over copper or coaxial cables. Similarly in FTTB, ONU is installed at individual buildings. And finally at the FFTH solution, each user has an ONU device at home. [2] gives a comprehensive overview of the issues and future trends in FTTX solutions.

Passive optical networking (PON) is the preferred choice of FTTX implementations due to both practical and cost considerations. In this architecture, the traffic sent downstream from the OLT is broadcasted to every ONU by means of a passive optical power splitter. In the upstream direction, it is necessary to use multiple access techniques. There ara several proposals focusing on different options: Time division multiple access (TDMA), subcarrier multiple access (SCMA), wavelength division multiple access (WDMA), optical code division multiple access (OCDMA), and possible combinations of these. Among these, TDMA is the simplest technique and most probably will be implemented first.

Another debate or competition is on the protocol that will be used in the data link layer. One option is the Asynchronous Transfer Mode (ATM). Full Services Access Network (FSAN) group of International Telecommunications Union (ITU) is working for the development of different Passive Optical Network (PON) standards based on ATM (ATM PON (APON), Broadband PON (BPON), and Gigabit PON (GPON)). Second option is the use of Ethernet in the data link layer. Ethernet in the First Mile (EFM) task force is developing the IEEE 802.3ah

standard for Ethernet PON (EPON) [7]. The dominance and cost advantage of Ethernet, together with the advances in the passive optical networking, makes EPONs a promising solution for the last mile.

To sum up, the evolution in access networks is driven by an exponentially increasing capacity demand from both residential and business users. Optical access technologies will probably be the preferred solution, since they promise more than enough capacity at least for the foreseeable future.

2.2 Core Networks

Long haul networks, also called as the backbone networks, carry large volumes of aggregated traffic over inter-regional distances (1000 km or more) and are optimized for distance and speed. The large amount of optical fiber installations in the backbone network and the use of dense wavelength division multiplexing (DWDM) provide a huge capacity potential. WDM (Wavelength Division Multiplexing) technology allows multiple data channels to be transmitted simultaneously over a single optical fiber. Today, WDM transmission systems can be built having capacities on the order of terabits per second, using more than one hundred channels at 10 gigabits per second each. Together with the fact that an optical cable may have more than 100 fibers, WDM provides a virtually unlimited capacity.

As the traffic volume has been increasing, optical fibers were first used for increasing capacity of point to point transmission. Electronic signals are converted to optical ones at one end of the fiber and back conversion is done at the other end. This phase is known as the first generation DWDM.

As point-to-point systems proliferate, wavelength channels are extended across multiple hops to maintain optical transparency as much as possible. At each network node, transit traffic is bypassed in the optical domain. Only the optical

signal destined to the node itself is converted to the electronic signal. Thus, unnecessary electronic to optical conversions are eliminated, which relieve the bottleneck of electronic processing. In the so called second-generation DWDM, the wavelength channels which are to be bypassed and which are to be added/dropped are fixed as a part of the network design. In the third generation of DWDM, the add/drop wavelength channels are reconfigurable which provides an opportunity for traffic engineering. Some of the important research subjects related to core networks include, optical signal amplification and regeneration, logical topology design and wavelength conversion issues. It is clear that the trend is towards an all optical transport network where optical wavelengths are transparently switched between network nodes. In that respect, control plane related issues and protocols (e.g. generalized multi-protocol label switching (GMPLS)) are popular topics of research.

The next generation of the backbone is expected to be based on optical burst switching (OBS) and optical packet switching (OPS) [27, 28, 29]. The basic idea of these approaches is to decrease the granularity of switching and hence increase the multiplexing gain. As mentioned, in second and third generation DWDM networks, wavelength channels are routed in the network. In OPS, optical packets are processed and routed individually which is similar to the routing process of packets in an electronic network. This approach requires optical processors and buffers in order to process and store optical packets but the current optical technology is not mature enough to build such components. Therefore, OBS is introduced as an intermediate step. In OBS, a burst of packets is sent and routed as a single entity. Control signalling and packet header processing is done in the electronic domain. Furthermore, OBS requires minimum buffering, which can be obtained by fiber delay lines available today. There is a growing research interest in the area of OBS/OPS and experimental testbeds are being developed to demonstrate the viability and efficiency of these approaches.

2.3 Metro Networks

Metro networks lie in between the access and backbone networks and covers the range from 20 km (suburban loops) up to 500 km (regional rings) [4]. Metro networks are further divided into two domains: metro access and metro core.

Metro access is also called as collector ring or metro edge and spans distance of 20-65 km. The traffic from the last mile networks and business are collected via distribution networks and aggregated at the central offices (COs). Metro access network connects these COs to each other and to the metro core through hub COs. Traffic in the metro edge has a hubbed traffic pattern and rings are natural choices of implementation in this part of the network.

Metro core (regional network, interoffice/feeder ring), in turn provides the connectivity between the hub COs and to the long haul backbone. Metro cores may extend up to 500 km and has mesh connectivity due to the any to any nature of the underlying traffic. Legacy infrastructure in the metro is based on time division multiplexing (TDM) architecture and optimized for circuit oriented services such as telephony. SONET/SDH is used both in metro access and metro core. OC-3 (155 Mbps) and OC-12 (622 Mbps) are commonly used in metro access part of the network. In the metro core, virtual SONET/SDH rings over mesh connected nodes are constructed with OC-48 (2.5 Gbps) and OC-192 (10 Gbps) speeds.

As discussed in Section 2.1, last-mile networks have been adapted to the increase of the packet based traffic demand with new technologies and increased bandwidth capacities. With DWDM installations, long-haul core is already capable of carrying larger volumes of traffic. On the metro arena specialized intermediate protocol layers and overlays are used on top of the TDM architecture, such as asynchronous transfer mode (ATM) and frame relay (FR). But with their scalability, cost and operational complexity problems, these solutions are far from being efficient for the data traffic. Hence, new solutions are required to overcome these limitations. Metro core networks are now experiencing an evolution similar to the backbone network. That is, the conventional SONET/SDH based metro

core is being replaced by WDM based solutions. The first step is the utilization of point-to-point fiber reliefs, then static add/drop rings and finally reconfigurable add/drop rings are evolving. However, metro access part of the network has been lagging behind these developments. With its legacy TDM architecture, low bandwidth scalability, long provisioning cycles and data port inefficiency, it became a bottleneck between the high speed last-mile networks and long-haul core. This bottleneck effect is called as the "metro gap".

There is a growing interest and research on improving the metro access performance. In addition to increasing the capacity, the new architecture should also be compatible with the specific requirements of metro networks. Among these, cost effectiveness is the primary one due to smaller number of customers served. Besides, low levels of flow aggregation results in rapidly changing traffic patterns which makes dynamic reconfigurability an important issue. New solutions should also address scalability and multi-service, multi-protocol (transparency) requirements due to the variety of last mile technologies in use. There are several emerging technologies for the metro access area as discussed next.

2.3.1 Next Generation SONET/SDH (NGS)

To overcome the limitations of legacy infrastructure, SONET/SDH is tailored to carry data traffic more effectively [30, 31, 32]. One major shortcoming of the TDM architecture is the inefficiency of mapping data traffic to SONET channels. For example, Gigabit Ethernet requires full 2.5 Gbps OC-48/STM-16. To alleviate this problem, Virtual Concatenation (VCAT) is developed. With VCAT it is possible to combine multiple smaller tributaries into a VC group (VCG) to better match non-TDM demands. To meet the dynamic reconfigurability needs, link capacity adjustment scheme (LCAS) protocol is defined as an NGS addition, so that the number of assigned VC trials can be re-adjusted on the fly. LCAS can also be defined as a low-level bandwidth capacity control protocol for virtual concatenation. Furthermore, for transparency requirements generic framing procedure (GFP) is developed [33]. It enables the mapping of diverse protocols onto byte-synchronous TDM channels.

The most important advantage of NGS, is the leveraging of the existing infrastructure. It also supports gradual deployment without costly changes to management systems. But it is obviously not the most efficient way to carry packet based traffic, due to its data-over-circuit approach and complexity.

2.3.2 Resilient Packet Ring (RPR)

Resilient Packet Ring (RPR, IEEE 802.17) is a packet networking technology which combines the best features of SONET/SDH (simplified connectivity, resiliency) and Ethernet (low cost, statistical multiplexing) [34]. Like SONET/SDH it uses a fiber ring topology but it is optimized to carry packet traffic. It has resiliency properties comparable to SONET/SDH and supports multiple services ranging from simple data traffic to latency/jitter sensitive traffic such as voice and video. With its spatial reuse protocol (SRP) bandwidth is only consumed between the source and destination nodes across the ring, which increase the efficiency. It uses a modified Ethernet medium access protocol (MAC) which inherently supports broadcast and improves fairness [35]. On the downside, it lacks standardized support for legacy TDM voice and leased line services.

2.3.3 WDM Based Solutions

As discussed in Sections 2.2 and 2.3, WDM is used intensively in backbone and regional networks. The steady growth of end user data rates and the desire for supplying gigabit per second capacity for high end customers make WDM a promising solution for metro access networks as well. Since the metro core and backbone networks rely on WDM, this approach also has the advantage of compatibility with the higher levels of the network. It is also capable of hosting various infrastructures transparently, such as legacy TDM, NGS and RPR. To a large extent, the barrier in front of the deployment of WDM in metro access networks has been the cost, but as the technology matures prices of optical components fall. Moreover, shorter spans required in these networks permit the use of passive (un-amplified) transport which eliminates the need for expensive

devices. Finally, the required number of wavelengths is less compared to a core network which makes it possible to use coarse WDM (CWDM) instead of DWDM to further decrease the cost. Several projects are being developed to build and demonstrate an efficient metro access network based on WDM technology.

Optical Regional Access Network (ORAN) Project [8, 9], proposes a flexible wavelength routed WDM network. A ring topology is preferred. It uses 64 wavelengths per fiber, and the number of fibers is between 2 and 30. Distribution networks are connected to access nodes and the ring connects these access nodes to each other and to the backbone through Egress Nodes. It also proposes a totally passive distribution network which can deliver WDM all the way to the end user.

A similar architecture is being developed by MIT Lincoln Laboratory with the name Next Generation Internet-Optical Network for Regional Access with Multiwavelength Protocols (NGI-ONRAMP) [10, 11]. 10-20 access nodes and 20-100 users per distribution network are anticipated. The number of wavelength channels is between 10 and 100, each at data rates 2.5 Gbps, 10 Gbps or higher. Like ORAN project, distribution networks are passive and can carry WDM channels as well as local distribution wavelength channels.

There are also packet over WDM approaches for the metro networks. One of them is the Hybrid Opto-Electronic Ring Network (HORNET) project [12], developed by the Stanford University and Sprint. It is argued that with the increase in the peer-to-peer communications, the hubbed traffic pattern of the metro access network will shift towards an any to any pattern. Hence, HORNET architecture is designed to enable all nodes to communicate more directly with other nodes, as in a meshed network. In this architecture, each node on the ring has a fixed wavelength optical receiver and a fast tunable transmitter. The sender node tunes its transmitter to the wavelength that the intended receiver is tuned. A smart node architecture, and a medium access control (MAC) protocol based on carrier sense multiple access with collision avoidance (CSMA-CA) is also developed to support the operation of this network.

Another example for packet over WDM ring metro network is the Ring Optical

Network (RingO) project [13] carried out by a consortium of Italian universities. It implements the control plane in electronic domain while the transmitted data is kept in optical domain. The general architecture is similar to the HORNET. The number of wavelengths is equal to the number of nodes in the ring and each node is tuned to receive a different wavelength channel. There is a tunable transmitter at each node which is used to transmit at the wavelength allocated to the receiving node. In this work a node structure and a MAC protocol are also proposed and an experimental testbed is developed for demonstration purposes.

Finally, the European Information Society Technologies (IST) funded project, Data and Voice Integration over DWDM (DAVID) [14] aims at proposing a viable approach toward Optical Packet Switching by developing networking concepts and technologies for future optical networks. It covers both the metro and wide area networks. The metro network has a ring topology consisting of one or more fibers operated in DWDM regime. Each wavelength is used to transport optical packets of fixed duration in time. A MAC protocol is developed to select the wavelength and time-slot to be used for transmission so that the optical path is kept bufferless. The header processing is still done in electronics while the payload is switched transparently in the optical domain. The WDM rings are interconnected to other rings via a bufferless hub which also controls the resources.

To sum up, metro access networks are going through a rapid evolution phase to catch up with the developments in the rest of the communication architecture. In this process, reconfigurability is seen as a major requirement. In the following chapter, wavelength routed IP/WDM metro access network architecture is discussed in more detail and dynamic reconfiguration problem is introduced.

Chapter 3

IP/WDM Metro Access Networks and the DWA Problem

As discussed in Chapter 2, IP/WDM is a promising architecture for future metro access networks. Dynamic reconfigurability feature of IP/WDM enables traffic engineering and efficient resource utilization. In this chapter, a basic overview of the operational principles of the IP/WDM metro networks and issues related to bandwidth allocation are discussed. The DWA problem is introduced along with the modeling assumptions and performance measures. Related literature review is also presented in this chapter.

3.1 IP/WDM Metro Access Network Architecture

The IP/WDM access network architecture considered is similar to the one defined by the NGI ONRAMP consortium [10, 11]. ONRAMPs are proposed as high speed optical metropolitan and small regional area networks which are low cost and easy to provision and manage. IP data is routed directly over the WDM physical layer (IP over WDM) and intermediate protocol layers, such as ATM

and SONET are eliminated. With IP/WDM a degree of transparency is achieved which enables the network to support heterogeneous traffic with different bit rates and data formats.

A high level view of an IP/WDM access network is shown in Figure 3.1. It consists of a single feeder ring which connects access nodes (ANs) to each other and to the backbone network. Each AN is located at a CO and is used to connect high-speed customers and distribution networks to the feeder ring. That is, each access node serves as a gateway to distribution networks on which the end users reside. The feeder ring is also connected to a backbone network via a hub (gateway) node located at the Hub CO. The ring may be built using a single fiber or multiple fibers, where each fiber supports tens of wavelengths. The traffic from distribution networks are aggregated at the corresponding ANs and transported to the hub node on multiple lightpaths, which are individually allocated wavelengths. Finally, the hub node forwards the traffic to the backbone network. For the downlink case the traffic follows the reverse path. It is envisioned that there may be 10 to 20 nodes in the feeder ring and 20-100 users on each distribution network. The feeder ring can carry 10 to 100 wavelengths channels at data rates of 2.5 Gbps (OC-48), 10 Gbps (OC-192), and potentially higher. To maintain connectivity between the hub node and ANs, each AN must be allocated at least one wavelength channel all the time. This implies that the number of wavelengths is greater than the number of ANs in a wavelength switched IP/WDM ring network since each node is allocated a separate set of wavelengths.

The hub node is responsible for the resource management of the ring. It allocates separate wavelength channels to ANs and the logical topology of the network, i.e., lightpaths between ANs and the hub node, can be changed by dynamically assigning wavelengths. This feature enables both dynamic provisioning of the network resources and reconfiguration of the network topology. Dynamic provisioning is the allocation of network resources to a user as needed for a limited period of time. On the other hand, reconfiguration of the logical topology is done to optimize the network performance as a whole. When some links of the network become congested, wavelengths can be reallocated to construct additional lightpaths to increase the capacity of these links.

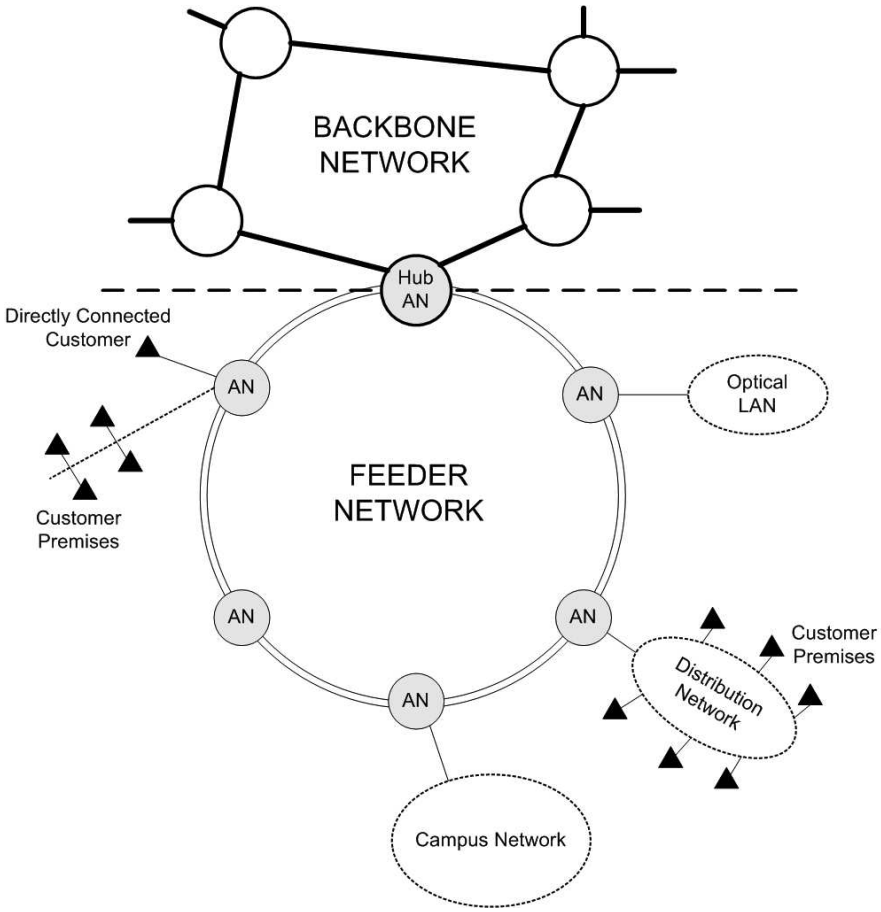


Figure 3.1: IP/WDM network architecture.

Each access node consists of an Optical Add Drop Multiplexer (OADM) and an IP router. The OADM selects the wavelengths to be added or dropped at a node and other wavelengths are routed all-optically in the feeder ring. Hence, IP traffic is transported from an access node to the hub node without passing through intermediate IP routers. The AN is also equipped with tunable optical receivers and transmitter so that wavelengths assigned to ANs can be changed dynamically by tuning these transmitters and receivers to support DWA.

The physical topology of the IP/WDM network is a ring. However, separate lightpaths are established between ANs and the hub node and these lightpaths are transparently forwarded at intermediate nodes. As a result, the logical topology seen by the upper protocol layers becomes a tree network and the hub node can be modeled as a simple multiplexer/demultiplexer as illustrated in Figure 3.2.

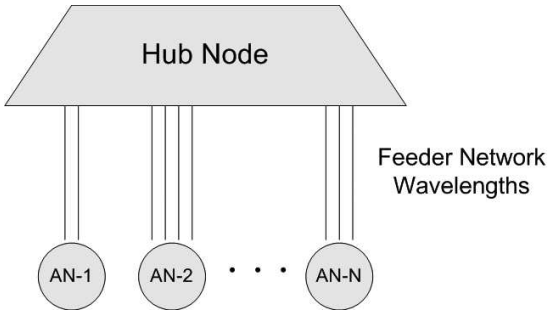


Figure 3.2: Logical view of the IP/WDM access network.

3.2 Resource Allocation in IP/WDM Metro Access Networks

The basic resource in a network is the transmission capacity. In an IP/WDM access ring transmission capacity is divided between ANs by assigning individual wavelength channels to each AN. Therefore resource granularity is the bandwidth of a wavelength channel and resource allocation corresponds to allocation of wavelengths to ANs. Basically, two strategies can be considered for resource allocation: static allocation and dynamic allocation.

In the static resource allocation scheme, capacity demand is measured over a period of time and projections are obtained. Considering the measurement errors and possible variations a safety margin is also calculated. Based on these data, resources are allocated and configuration is not changed in time. For the wavelength switched metro access ring, this approach may be called as *Static Wavelength Allocation (SWA)* since wavelengths assigned to access nodes are fixed as a part of the network design. For instance, if all the nodes have the same expected offered load, then the wavelengths should be evenly distributed between nodes. Static resource allocation approach has the advantages of simplicity and ease of management. But it requires the demands to have stable and predictable patterns. This condition holds for backbone networks where traffic is highly multiplexed and groomed, and experiences less variability. Therefore, static design of the logical topology and over-provisioning the capacity to handle measurement inaccuracies and traffic uncertainties is a commonly used approach. However, a

metro access network differs from a long-haul or backbone network in several key aspects. At the periphery of a metro access network, the level of demand granularity is often an individual traffic stream with its highly variable characteristics. This fact may cause large traffic deviations in very short time scales comparable with the duration of flows. The mobility of users in the area served by a metro access network also causes traffic variations on the order of hours. There may be periodical demand shifts as the density of customers in residential and industrial areas change during the day. Therefore SWA may result in inefficient usage of available capacity and unfair service delivery.

Dynamic resource allocation allows the resources to be dynamically configured to follow demand fluctuations. For the metro access ring this approach corresponds to *Dynamic Wavelength Allocation (DWA)* where the traffic at each node is monitored and the number of wavelengths assigned to ANs are changed accordingly. This scheme has the potential to improve the efficiency and fairness in capacity utilization compared to SWA. However, DWA has an overhead due to signalling requirements, reconfiguration of OADMs, and tuning latencies of the transmitters and receivers. As a result the wavelength channels being reconfigured becomes unavailable for a certain duration of time which is called as the *reconfiguration delay*, and denoted by τ. This corresponds to a loss of capacity and presents a trade-off between the reconfiguration costs and the responsiveness of the network to demand changes. Clearly, reconfigurations should not be very frequent, since unnecessary wavelength transfers between nodes decrease the capacity of the network and adversely affect the performance. Hence, it is desirable to minimize the number of network reconfigurations. However, postponing a necessary reconfiguration also has negative effects on the overall performance, since the network does not operate at an optimal point in terms of load balancing. Similarly, if the decisions are made solely by considering load balancing, even small changes in the traffic demands can lead to reconfigurations which may cause a significant decrease in network performance. Consequently, it is important to capture the trade-offs in an appropriate manner and allow their simultaneous optimization.

To illustrate the possible advantages of DWA consider a sample metro access

network consisting of 48 wavelength channels and 6 nodes, where 3 of the nodes (AN1, AN2, AN3) serve to residential areas, 2 nodes (AN4, AN5) are connected to business customers and the last node (AN6) serves as an access node for a campus network. Assume that the bandwidth demands at each node changes during the day according to Table 3.1, where the values are normalized with respect to the bandwidth of a single wavelength channel. The traffic demand between 00–08 is assumed to be negligible.

Table 3.1: Bandwidth demands at each AN as a function of time.

time of day (h)	**AN1**	**AN2**	**AN3**	**AN4**	**AN5**	**AN6**
08–12	1.6	1.6	1.6	12	12	9.6
12–16	1.6	1.6	1.6	9.6	9.6	14.4
16–20	6.4	6.4	6.4	4.8	4.8	9.6
20–24	9.6	9.6	9.6	2.4	2.4	4.8
Avg	4.8	4.8	4.8	7.2	7.2	9.6

In the SWA scheme the wavelengths are distributed to ANs proportional to the average traffic demand. The resulting number of wavelengths allocated to each node are 6, 6, 6, 9, 9, 12, for ANs 1 to 6, respectively. With this allocation, capacity utilization levels are given in Table 3.2. First, it is observed that there are time periods during which the utilization level for an AN is greater than 1, meaning that the demand exceeds the maximum service rate. Thus, the system becomes overloaded with unserviced traffic as time progresses, and SWA can not produce a feasible wavelength allocation in this case. As a second observation, the capacity utilization levels show considerable variations which indicates an unfair service distribution.

Table 3.2: Capacity utilization levels for the SWA.

time of day (h)	**AN1**	**AN2**	**AN3**	**AN4**	**AN5**	**AN6**
08–12	0.27	0.27	0.27	1.33	1.33	0.80
12–16	0.27	0.27	0.27	1.07	1.07	1.20
16–20	1.07	1.07	1.07	0.53	0.53	0.80
20–24	1.60	1.60	1.60	0.27	0.27	0.40

With DWA, it is possible to reconfigure the wavelength allocation to match the demand at different time periods. Number of channels assigned to each node

with a sample DWA allocation policy is shown in Table 3.3. It may be verified that the utilization level at each AN for all time periods is 0.8.

Table 3.3: Wavelength allocation with DWA.

time of day (h)	**AN1**	**AN2**	**AN3**	**AN4**	**AN5**	**AN6**
08–12	2	2	2	15	15	12
12–16	2	2	2	12	12	18
16–20	8	8	8	6	6	12
20–24	12	12	12	3	3	6

To obtain a feasible wavelength allocation with maximum utilization level of 0.8, SWA needs to consider peak traffic rates and allocate 12, 12, 12, 15, 15, 18 wavelengths to ANs 1 to 6, respectively. Hence, with SWA 84 wavelengths, i.e., 75% more than used in DWA, are required and the average capacity utilization falls to 0.46.

This example demonstrates the potential improvements that can be achieved with DWA under non-stationary traffic. DWA is also expected to improve the performance even for the case of stationary traffic because of two reasons. First, SWA may result in allocations that may not be exact multiples of resource granularity which is equal to the bandwidth of a single wavelength channel. A simple example is a network with three wavelengths and two nodes with equal demand. With SWA one node is allocated 1 wavelength and the other node is allocated 2 wavelengths. However, DWA may result in an allocation of 1.5 wavelengths to each node when averaged over time. Second, the demand is stochastic and may exhibit fluctuations around the average value in short time scales. With DWA, wavelength allocation can be changed to follow these variations which may improve the performance as a result of statistical multiplexing.

3.3 DWA Mechanism and Reconfiguration Delay

The wavelength routed IP/WDM network is a centralized architecture since the lightpaths are established between the hub node and ANs. Therefore, the hub node can easily monitor the traffic on the ring and is responsible for the management and reconfiguration of capacity allocated to each AN. The number of flows destined to or originated at each access node can be counted at the hub node using one of the several techniques available in the literature (e.g., [36, 37, 38]). Based on this measurement, hub node may decide to initiate a wavelength switching action. The reconfiguration process requires the transmission of necessary signalling messages, processing of these messages at the nodes, and finally actual tuning of the transceivers at the nodes. The mechanism is slightly different for the downlink and uplink traffic cases as explained below.

For the downlink case, the optical transmitter and receiver are located at the hub node and AN, respectively. The outline of the steps required to switch the wavelength l from node i to node j may be as follows:

1. Hub node stops transmitting on wavelength l.
2. Hub node waits at least for $2 \times RTT$ to allow packets already transmitted on wavelength l to reach to node i.
3. Hub node sends a message to node i to stop reception on l and another message to node j to start receiving using wavelength l.
4. Node j tunes its optical receiver to l.
5. Node j send acknowledgement to hub node to inform the completion of the tuning.
6. Hub node begins transmission to node j on l.

Round-trip time (RTT) is the time required to transmit a message from the

hub node to an AN and back again, and for this case it is equal to the one ring-traversal latency. The total reconfiguration time is $4 \times RTT + T_{rx} + T_p$, where T_{rx} is the tuning delay of the optical receiver and T_p corresponds to the sum of time required for processing messages at nodes and generating low level hardware instructions.

For the uplink case, the optical transmitter and receiver are located at the AN and the hub node, respectively. Following steps should be performed to switch the wavelength l from node i to node j:

1. Hub node sends a message to node i to stop transmission on wavelength l.
2. Hub node waits at least for $2 \times RTT$ to allow packets already transmitted on wavelength l to reach to itself.
3. Hub node sends a message to node j to tune its transmitter to l.
4. Node j tunes its optical transmitter to l.
5. Hub node begins reception from node j on l.

The reconfiguration delay for this case sums up to $3 \times RTT + T_{tx} + T_p$, where T_{tx} corresponds to the tuning time of the optical receiver at node j.

For an IP/WDM ring of length 90 km, RTT is approximately 0.3 ms. The tuning times of the optical transmitters and receivers depend on the technology used. The processing time T_p is determined by the processor and software used in the OADM. The tuning times of optical transmitters and receivers depend on the technology being used and approximate values are as follows [39]. For the mechanically-tuned lasers T_{tx} is on the order of milliseconds. For acoustooptically- and electrooptically tuned lasers, tuning times are approximately 10 μs and 10 ns, respectively. Sub-nanosecond tuning times are also feasible with the injection-current-tuned lasers. The tuning times of optical filters determines the T_{rx} and with mechanically tuned filters, T_{rx} is on the order of tens of milliseconds. The tuning time of MZ chains is also on the order of milliseconds. T_{rx} is about 10 μs and several nanoseconds for the acoustooptic

and electrooptic filters, respectively. For the liquid-crystal Fabry-Perot Filters the tuning times are also on the order of microseconds. Fast tunable lasers are generally preferred in packet IP/WDM, where the optical transmitter should be able to change frequencies between consecutive packet transmissions. For traffic engineering purposes, slower transmitters and receivers may be preferred due to their cost advantages. T_p may also be expected to be in the order of milliseconds or 10 milliseconds depending on the hardware being used. As a result, the total reconfiguration delay, τ, is expected to be on the order of 10 milliseconds.

In this work, the average reconfiguration delay is set to 50 ms as a conservative value. Smaller values of reconfiguration delay further increases the benefits of DWA since the cost of reconfiguration decreases in this case. The effects of the value of reconfiguration delay on the performance are analyzed in Section 6.4.2, where the average delay is changed in the range of 0 to 1 s.

3.4 DWA Framework

In order to be able to develop solutions for the DWA problem, an appropriate framework is constructed. This framework includes the modeling assumptions, performance metrics and a formal definition of the DWA problem as described in the following subsections.

3.4.1 Performance Metrics

To evaluate the effectiveness of wavelength allocation strategies quantitative measures are needed. Throughput and fairness are among the basic performance metrics for a network. In this thesis, *slowdown* and *holding cost* are used to measure throughput efficiency and *Jain's Fairness Index* is utilized to assess the fairness performance. These metrics are calculated based on flow level statistics as described in the following subsections.

3.4.1.1 Holding Cost

In resource allocation and queueing problems, it is generally assumed that each job waiting or being serviced in the system incurs a holding cost per unit time depending on the relative importance of the job class [18, 19, 20]. For the DWA problem, jobs correspond to flows in the network and each flow has the same relative importance. Without loss of generality, the cost per unit time can be taken as unity. Then, the total holding cost is simply equal to the sum of flow completion times. In this work, average holding cost is used as a throughput metric and defined as

$$HC_{avg} = \frac{1}{K} \sum_{k=1}^{K} FCT_k,$$

where K is the total number of flows and FCT_k is the flow completion time for the k^{th} flow.

With this definition, minimum cost is achieved when the average flow completion time is minimized or equivalently when the average throughput is maximized. Therefore, average holding cost can be used as a direct measure of average flow throughput.

3.4.1.2 Slowdown

In addition to increasing average throughput, it may be desirable to balance the throughput achieved by each flow. This corresponds to having flow completion times that are proportional to flow sizes. Holding cost is not a sufficient measure for this purpose since it considers only the average value of the throughput. Therefore, slowdown is used as a normalized throughput metric and defined as the ratio of actual flow duration to the time that would be required if the flow had dedicated access to an entire wavelength channel [24]. So, a slowdown value larger than 1 indicates that only a fraction of a channel is used to transmit the flow. The average slowdown experienced by all flows is used to evaluate the throughput efficiency of the network. Slowdown for the k^{th} flow, SD_k, can be expressed as

$$SD_k = \frac{FCT_k}{FS_k/B}$$

where FCT_k and FS_k are the completion time and the size of the k^{th} flow, respectively, and B denotes the capacity of a single wavelength channel. The average slowdown, SD_{avg} is defined as

$$SD_{avg} = \frac{1}{K}\sum_{k=1}^{K} SD_k = \frac{B}{K}\sum_{k=1}^{K} \frac{1}{FS_k} FCT_k, \tag{3.1}$$

where K is the total number of flows. Slowdown can also be interpreted as a holding cost where the cost associated with each flow is inversely proportional to its size which implies that short flows should be served in less time compared with long flows, to decrease the cost.

Slowdown and holding cost metrics can be compared on a simple example. Assume that there are two flows with sizes B and $4B$ bits and consider two possible cases given in Tables 3.4 and 3.5. For both cases, average throughput is $0.5B$ bps and the average holding cost is 5. However, average slowdown is found to be 3.125 and 2 for the first and second cases, respectively. It is seen that the average slowdown is lower for the second case where the throughput achieved by the flows have equal values.

Table 3.4: Comparison of slowdown and holding cost metrics - Case 1.

Flow No	**FS (bits)**	**FCT (s)**	**Throughput**	**Slowdown**
1	B	5	0.2B	5
2	4B	5	0.8B	1.25

Table 3.5: Comparison of slowdown and holding cost metrics - Case 2.

Flow No	**FS (bits)**	**FCT (s)**	**Throughput**	**Slowdown**
1	B	2	0.5B	2
2	4B	8	0.5B	2

3.4.1.3 Fairness Index

Service fairness is another important issue in resource allocation. It may be desirable that resources are shared between each job in the system evenly. For the DWA problem, fairness is related to the discrepancies between slowdown

values of different flows. To be able to compare wavelength allocation methods, a quantitative measure of fairness is required and Fairness Index defined in [40] is used for this purpose. Fairness index can be adapted to the DWA problem as

$$FI = \frac{(\sum_{k=1}^{K} SD_k)^2}{K \sum_{k=1}^{K} SD_k^2},$$

where K is the number of flows and SD_k denotes the slowdown corresponding to k^{th} flow as defined in 3.1.

Fairness index has several useful features. First, it is independent of scale since the unit of measurement does not matter. Second, it is bounded between 0 and 1. A totally fair system has an index of 1 and a totally unfair system has a fairness of 0. Hence, it is easy to interpret and compare fairness levels for different cases. Moreover, the index is continuous and any slight change in allocation shows up in the fairness index.

3.4.2 Assumptions

For the sake of tractability, several assumptions and simplifications are done in the formulation of the DWA problem.

First of all, it is assumed that the local traffic between ANs is negligible and this traffic can be relayed through the hub node. This is a reasonable assumption since observations within access networks have shown that approximately 90% of data traffic is originated at or destined for points outside the network [24]. However, for the cases where this assumption does not hold, it may be more suitable to operate IP/WDM in packet mode which corresponds to another architecture conceptually.

To simplify the problem, only the traffic from ANs to hub node is considered. However, this is not a restrictive assumption because the uplink traffic (traffic from ANs to the hub node) and the downlink traffic (traffic from the hub node to the ANs) are transmitted on independent set of wavelengths. Hence, a similar formulation can be used for the traffic in the reverse direction.

The problem definition inherently assumes that the flow throughputs are restricted by the metro access network. This is a realistic assumption for practical networks where the capacity of backbone networks are much larger compared to metro access networks and high speed users are directly connected to ANs. When the bottleneck is not the metro access network or with multiple bottlenecks, the problem gets more complicated.

It is assumed that reconfigurations are initiated at flow arrival and departure instants. Intuitively, it is at those instants that reconfiguration may become advantageous, and if they do, they should be performed without delay [41].

Flow arrivals to each node is modeled by an independent Poisson process and flow sizes are exponentially distributed. The reconfiguration delay is a random variable which also has an exponential distribution. Although these assumptions are not valid in general, they are used for mathematical tractability and easy interpretation of results. However, it is possible to extend the problem to incorporate more realistic traffic arrival and flow size statistics as a future research. For instance, flow arrivals can be modeled by a Markov Modulated Poisson Process and flow sizes can have a general distribution such as Pareto distribution. The methods developed in this thesis can then be adapted to these processes with the use of a semi-Markov formulation.

It is also assumed that a packet scheduler is used at each AN. Therefore, each of the flows at a node uses a fair share of the bandwidth available at that node. This also implies that a flow may use the capacity of multiple wavelengths allocated to the corresponding node. With this assumption, the total bandwidth allocated to a node can be simply treated as a single channel with an aggregated capacity of all channels.

Finally, flows are assumed to be elastic so that they adapt their data rates to the available capacity and they do not have any peak rate. Therefore, flows are able to use the available capacity in an efficient way. For real life TCP flows, this assumption is only partially correct as it will be analyzed in Section 8.2. The modification of the methods to incorporate this non-ideal TCP behavior is identified as a future research topic.

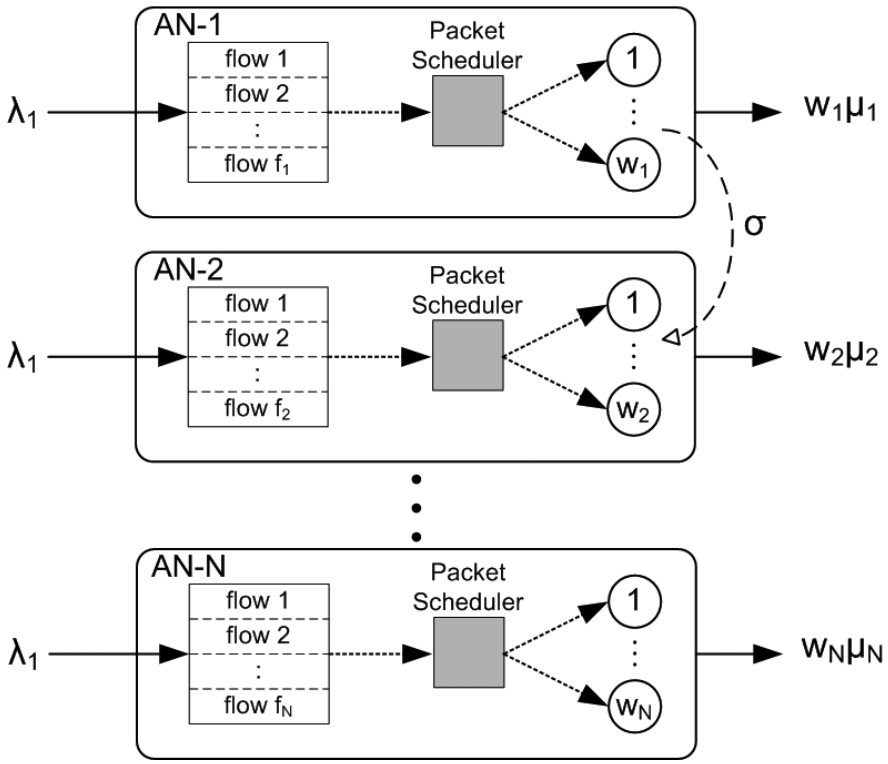


Figure 3.3: Network model used for the DWA problem.

3.4.3 Network Model

Using the assumptions of the previous section, the IP/WDM network can be modeled as a set of parallel queues served by a group of servers as illustrated in Figure 3.3.

The network has N access nodes which can be represented as N parallel queues. To each node i, flows arrive according to a Poisson process with rate λ_i. Active flows at each node correspond to the jobs at the related queue. Due to the presence of packet schedulers, the service discipline of queues can be approximated as processor sharing (PS), i.e., all of the flows at a node is served in parallel with equal rates.

Total number of wavelength channels is W, where $W > N$, and each channel corresponds to a single server in the queueing model. The capacity of a server is equal to the bandwidth of an individual wavelength channel, B. Total bandwidth

allocated to node i is the product of the number of channels assigned, w_i, and B. Flow sizes at node i are exponentially distributed with mean B/μ_i. Therefore, departure rate of flows at node i is $w_i \times \mu_i$.

Each flow arrival or departure event constitutes a decision epoch at which a single wavelength can be switched between nodes as long as there is no wavelength currently being moved. This action causes the wavelength being switched to be unavailable during a period of time called *reconfiguration period*. The length of the reconfiguration period is equal to the reconfiguration delay which is denoted as τ. It is assumed that τ is exponentially distributed with mean $1/\sigma$. In the queueing model, wavelength switching action corresponds to moving a server from a queue and assigning it to another queue. And during the reconfiguration delay the server remains idle.

3.4.4 Problem Definition

In general terms, DWA problem can be defined as the maximization of network efficiency by dynamically reconfiguring wavelength assigned to ANs. More specifically, the goal of DWA is taken to be twofold. The primary objective is to minimize the average slowdown experienced by the flows in the network. As a secondary objective, the maximization of fairness is desired.

To simplify the problem, reconfiguration is allowed only at flow arrival or departure instants, i.e., when the number of flows at any AN is changed. Therefore, each flow arrival or departure event is called as a decision instant. In addition, at most one wavelength is allowed to be in the switching state at any time. This also indicates that reconfigurations are limited to single switching actions. Hence, at most one wavelength can be switched between ANs at each decision instant if no wavelength is being switched already.

As a result, the DWA problem is defined as follows:

Minimize average slowdown and maximize fairness by reconfiguring the wavelength allocation subject to the following constraints:

1. *Decision instant constraint:* Reconfiguration can be performed only at flow arrival or departure instants.

2. *Single switch constraint:* At most one wavelength can be in the switching state at any time.

3. *Reconfiguration delay constraint:* When a wavelength is decided to be switched between ANs, it stays in the switching state for the duration of reconfiguration delay before being actually allocated to the new node.

4. *Wavelength unavailability constraint:* A wavelength which is in the switching state can not be used to carry traffic.

5. *Connectivity constraint:* At least one wavelength must be allocated to each AN at any time.

The solution of the DWA problem results in the optimum switching actions corresponding to each possible network state, where the state of the network is determined by the number of flows and number of wavelengths at each AN, and whether a wavelength is already in the switching state. The action may be to switch a single wavelength between a pair of nodes or not to make any switching and keep the wavelength allocation as it is. Solution of the DWA problem can also be called as the optimum switching policy.

Obtaining an optimum switching policy is a challenging task mainly due to the presence of the switching cost, which manifests itself as the unavailability of the wavelength being switched for the duration of the reconfiguration delay. This delay introduces a trade-off between switching costs and traffic adaptability. Therefore, a switching action should be performed if the long term benefits overweighs the switching costs. Moreover, the single switch constraint necessitates the selection of the most beneficial action. Note that in the absence of switching costs and single switch constraint the DWA reduces to a simpler problem of determining best wavelength allocation given the number of flows at each node.

3.4.5 Optimum Static Bandwidth Allocation

Using the network model described in Section 3.4.3, it is possible to calculate the bandwidth that should be allocated to each node in order to maximize the average flow throughput. The following lemma gives the optimum amount of capacity, in units of the channel bandwidth to be assigned to node i

Lemma 1. *For static bandwidth allocation, the optimum amount of bandwidth that should be allocated to each node i in order to maximize the average flow throughput is given by*

$$\bar{w_i} = \frac{\sqrt{\lambda_i}}{\mu_i} \frac{(W - \Lambda)}{\sum_i \left(\sqrt{\lambda_i}/\mu_i\right)} + \lambda_i.$$

where $\Lambda = \sum_i \lambda_i$.

Proof. Maximum average throughput is achieved when the average flow completion time is minimized. The first order statistics for the M/M/1-PS queue is same with M/M/1-FCFS queue [42]. In other words, although variance of the flow completion times differ the average flow completion times are equivalent for both service disciplines. Therefore, the expected flow duration at node i is given by the well known waiting time equation [43]

$$E[T_i] = \frac{1}{\bar{w_i}\mu_i - \lambda_i}.$$

The average expected flow duration for the network can be written as

$$T_{avg} = \sum_{i=1}^{N} \frac{\lambda_i}{\Lambda} \frac{1}{\bar{w_i}\mu_i - \lambda_i}.$$

The optimum static wavelength allocation is obtained by minimizing T_{avg} subject to the constraint that wavelengths allocated to nodes sum up to W, i.e.,

$$\sum_i \bar{w_i} = W. \tag{3.2}$$

The method of Lagrange Multipliers is used to calculate optimum bandwidth at each node as follows

$$\frac{\partial}{\partial \bar{w}_i}\left(T_{avg} + c\left(\sum_i \bar{w}_i - W\right)\right) = 0$$

$$-\frac{\lambda_i}{\Lambda}\frac{1}{\left(\bar{w}_i\mu_i - \lambda_i\right)^2} + c = 0$$

$$\bar{w}_i = \frac{1}{\mu_i}\frac{\sqrt{\lambda_i}}{\sqrt{c}\sqrt{\Lambda}} + \lambda_i \tag{3.3}$$

Using (3.2), c is found to be

$$c = \frac{\left(\sum_i \left(\sqrt{\lambda_i}/\mu_i\right)\right)^2}{\Lambda\left(W - \Lambda\right)^2}$$

Substituting c in (3.3) gives the desired result

$$\bar{w}_i = \frac{\sqrt{\lambda_i}}{\mu_i}\frac{(W - \Lambda)}{\sum_i \left(\sqrt{\lambda_i}/\mu_i\right)} + \lambda_i.$$

□

For $N = 3$, $W = 7$, $\mu_i = 1$ for all nodes i, and $\lambda_1 = \lambda$, $\lambda_2 = 2\lambda$ and $\lambda_3 = 4\lambda$, the optimum static allocation is plotted in Figure 3.4 as a function of λ.

It is observed that as λ approaches 1, fraction of bandwidth allocated to nodes become proportional to the arrival rates. It is also seen that the optimum bandwidth is not an integer multiple of a wavelength channel unless $\lambda = 1$. Therefore, static wavelength allocation is generally suboptimum when the wavelength channel granularity is taken into account. It may also be possible to calculate the optimum bandwidth allocation to minimize the average expected slowdown. However, due to its complexity it is not discussed in this thesis.

3.5 Related Work

Dynamic control of queueing systems has been a subject of great interest due to its potential applications in numerous areas including manufacturing, communication networks, multiprocessor systems, service grids, etc. In the simplest and

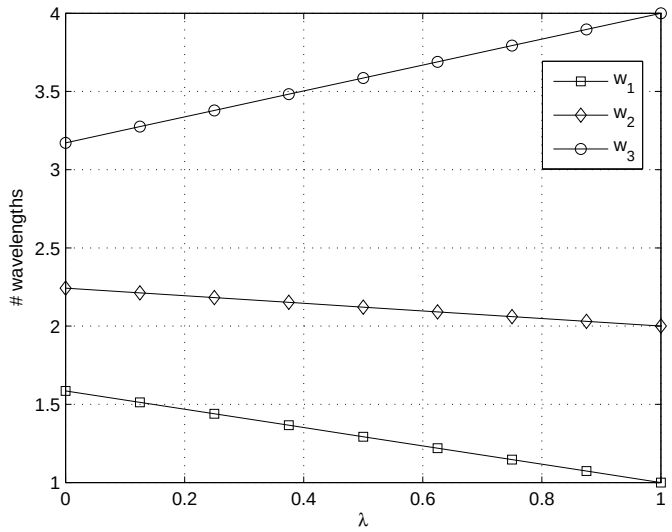


Figure 3.4: Optimum static capacity allocation for a 3-node test network.

one of the most studied model there are K parallel infinite buffer queues served by a single server [44]. The server switches between queues to process the jobs in different queues. There is no delay or cost associated with the switching action. The objective is to minimize the average weighted expected sum of queue lengths or of workload in the queues. This objective also corresponds to the minimization of the weighted holding cost. The weight factors are given by some positive constants $c_i, i = 1, \ldots, K$. The inter-arrival times have general distribution and service times in queue i are assumed to be exponentially distributed with parameter μ_i. It has been known that the optimal policy is the so called $c\mu$-rule [45]. It is a fixed priority policy where the different queues are ordered according to the decreasing order of the product of the weight c_i times the service rate μ_i, and a queue is served only if those queues with a higher product of $\mu_i c_i$ are empty. Such type of policy is called a *list policy*, where the policy has an associated list of the queues and processes the job whose queue is highest in the list. The results are also generalized to discrete time settings in [18, 46, 47].

If switching of the server incurs a time delay and/or cost then the system is called a *polling system.* The range of applications in which polling models can be used is very broad. Token passing local area networks with distributed channel access control and resource arbitration, and load sharing in multiprocessor

computers are just a few examples where polling systems are used for modeling and optimization [48] and there is an extensive literature on the subject (more than 200 references included in the survey paper [49]). Two control problems arise in the context of polling. The first problem is called schedule problem and is related to the order in which the server visits queues. The second problem is to determine the service duration at each queue, i.e., how many jobs are to be served upon visit to a queue. The former question corresponds to the routing policy whereas the latter question determines the service policy. A polling policy can be defined as a sequence of decision on whether to serve a customer, idle the server, or switch the server to another queue. The polling problem is studied for the case of two parallel queues in [50, 19, 20, 51]. [50] analyzes a special case of the polling model with $c_1\mu_1 = c_2\mu_2$ and show that the optimal policy serves each queue exhaustively and switch from an empty queue to the other if the number of jobs in that queue exceeds a certain value. [19] considers the case with switching costs and studies the optimal preemptive dynamic assignment of the servers to the queues. They conclude that it is unlikely to describe the optimal policy easily. Utilizing an MDP model, they obtain dynamic programming equations and with the asymptotic analysis of these equations they obtain approximate threshold policies. In [20], it is argued that the polling problem with two asymmetric queues appear to be analytically intractable and use a heavy traffic approximation to obtain threshold policies similar to [19]. They consider both cases of switching cost and switching time and conclude that each one leads to a fundamentally different qualitative solution. In [51] a specific polling policy is analyzed for both zero and nonzero switch-over times. For both cases they investigate ergodicity conditions, and derive exact expressions for the steady-state queue length and sojourn time distributions, and the joint queue length distribution at customer departure epochs. They also develop approximations for the mean queue lengths.

Polling systems with more than two queues are considered in [52, 53, 54]. Partial characterization of the optimum policy is obtained for homogeneous polling systems with switch-over times in [52]. [53] also partially characterize the optimum policy for the case of general service distributions and set up costs. They

also provide a simple heuristic scheduling policy. In [54] switch-over costs are considered and heuristic methods are developed based on the partial characteristics of the optimum policy.

Polling models with multiple servers have been considered by just a few papers. The analysis of a limited class of polling systems with multiple coupled servers is done in [55] and results on distribution of the waiting time, marginal queue length, and joint queue length at polling epochs are obtained. Stability issues for the multiple server case are studied in [56]. However, optimum policies for models with multiple servers and multiple queues have not been studied at all.

The DWA problem resembles a polling system with multiple coupled servers and multiple queues, where there is a switch-over time associated with reconfiguration actions. However, DWA problem has some specific features which have not been studied in the context of polling systems theory. First, the connectivity constraint necessitates the assignment of at least one server to each queue at all times and preemption of a queue is not allowed. Moreover, the literature on polling systems discusses queues which are operated on a First Come First Serve (FCFS) basis. On contrary, in DWA problem the queue discipline can be approximated as Processor Sharing (PS), since each flow at a node is served in a parallel fashion. As a result, obtaining optimum policies for the DWA problem in the context of polling systems is a challenging, if not impossible, task. Therefore, solution methods specifically designed for the DWA problem are required.

There are several works on the subject of reconfiguration in wavelength switched IP/WDM access networks. The network architecture considered in [24, 57] is very similar to the one assumed in this thesis. A flow level modeling is used in [24]. Flows arrive at each access node according to a Poisson distribution and the amount of data transmitted within a flow has a heavy tailed Pareto distribution. The objective of reconfiguration is to improve the network performance which is measured in terms of slowdown. It is argued that keeping all wavelengths in the system evenly loaded minimizes the mean slow down. Based on this idea a simple heuristic policy is proposed. The policy makes wavelength switches between nodes to improve the load balance. The simulation results obtained for

zero switching time case are compared with static wavelength allocation policy and it is shown that reconfiguration results in a significant decrease in slowdown and an apparent reduction in the tail of the slowdown distribution. The problem discussed in [24] is very similar to the DWA problem considered in this work. However, in this thesis a comprehensive model of the problem is developed, and an exact solution of the problem is introduced along with an efficient heuristic method. The method devised in [24] is also used to obtain numerical results and make performance comparisons.

In [57] the same problem is analyzed and two approximate queueing models are used to demonstrate potential benefits of reconfiguration and obtain performance bounds. Like [24] a stationary traffic model is used, however the analysis is based on a packet level model, where Poisson traffic is assumed to arrive at each node for the static traffic case and a two state Markov-modulated Poisson process (MMPP) is used for the dynamic traffic. Average packet delay is used as the performance metric. In the first queueing model, Continuous Bandwidth Model (CBM), a single server is used to model all of the wavelengths. In the second model, Wavelength Model, each wavelength is modeled as an independent server. During the system reconfiguration time, the servers in both models are assumed to be idle. Each access node is represented as an input queue and four different reconfiguration policies are discussed. In each case, a server works on a queue until it is empty then moves to the next queue determined by the reconfiguration policy. The policies are compared using simulations of the CBM model and it is concluded that reconfiguration can improve bandwidth utilization even in stationary traffic conditions under both symmetric and asymmetric average arrival rates.

Similar dynamic reconfiguration issues also arise in a single hop WDM packet ring, which is another promising architecture for future metro networks [25, 58]. In this network access nodes are connected to each other in a ring topology. Each node is equipped with a fast tunable optical transmitter, and a receiver which is tuned to a specific wavelength. When a node needs to send a packet, it simply tunes its transmitter to the destination nodes's receiver wavelength and establish a temporary single-hop connection for the duration of the packet transmission. A medium access control protocol is used to arbitrate transmissions

on the same wavelength, i.e., a time division multiple access (TDMA) scheme is used in each channel. It is generally envisioned that the number of wavelength channels is smaller than the number of nodes. Therefore, each channel is shared by multiple receivers and a decision problem arises concerning the allocation of different receivers to wavelength channels. If receivers are not tunable the allocation is permanent and can not be updated in response to changes in the traffic pattern. Then the problem is a static wavelength assignment problem. Alternatively, slowly tunable receivers can be used instead of fixed receivers so that the wavelength allocation can be changed dynamically to follow long term traffic variations. The performance measure is generally taken to be the average packet delay and the objective is to balance the traffic load on each channel by properly assigning receivers to wavelengths, which corresponds to minimizing the maximum load across all channels. The receivers can not receive packets during the tuning process which lasts a period of time equal to the tuning latency. Hence, there is a trade-off between load balance and reconfiguration costs, which translates into a dynamic reconfiguration problem.

A solution approach to this problem appears in [25]. In this work a mesh traffic pattern is assumed and reconfigurations are triggered with the changes in traffic demand, such as flow arrival and departures. Dynamic wavelength assignment is decomposed into the subproblems of "How to reconfigure" and "When to reconfigure". The former problem is related to obtaining load balance among channels with minimum receiver re-tunings and solved approximately using the Generalized Longest Processing Time (GLPT) algorithm given in [59]. The latter problem is to decide on whether to reconfigure or not at a given network state and solved using an MDP approach. In order to have a manageable sized state space, the state of the network is approximately represented with a feature called "degree of load balancing (DLB)" which measures the distance of the system to the ideally balanced situation. The cost of reconfiguration is related to the number of re-tunings required. Assuming that DLB changes over time according to a Markovian process, the MDP is solved to yield a threshold type policy. The policy gives the optimum action (reconfigure to the new wavelength assignment calculated by GLPT or keep the current configuration) as a function of DLB and

number of re-tunings required.

A similar problem is considered in [58], where the reconfiguration algorithm relies on measurements to detect the traffic pattern which is assumed to be unknown. The optimum way of reconfiguring receivers and deciding on whether to schedule the reconfiguration actions are again discussed as decoupled problems. A mixed integer linear programming (MILP) formulation is provided for the first problem, i.e., obtaining load balance among channels with minimum receiver re-tunings and it is shown to be NP-hard. And a heuristic approach is developed based on Longest Processing Time (LPT) and Maximum Weight Matching (MWM) algorithms. A simple solution is proposed for the second problem, where the wavelengths are reconfigured if the load of the maximum loaded channel improves more than a threshold percentage. The resulting reconfiguration policy is evaluated under two different traffic measurement schemes, namely when the traffic is measured at the access nodes and when the traffic is measured at a central node.

In [60] a simple approach is presented for reconfiguring the wavelengths. They propose a new algorithm named The Most and Least Loaded Channel Balance (MLLCB), where one of the receivers, which are assigned to the most loaded channel is exchanged with the appropriate receiver in the least loaded channel. Therefore, incremental reconfigurations are performed to balance the channel loads. The performance of the algorithm is compared with GLPT which is proposed in [59] and it is argued that MLLCB decreases the total number of retunings and results in a better average DLB value.

The DWA problem considered in this thesis differs from these studies in several aspects. First, the number of channels is assumed to be larger than the number of nodes and one or more separate wavelength channels are assigned to each node. The problems of "How to reconfigure" and "When to reconfigure" are jointly handled to produce solutions which aim to gradually improve the load balance in the network by making single wavelength switches at each decision instant in contrast to [25, 58] where the reconfiguration actions result in multiple switches to take the network into the most balanced state possible. The cost of

reconfiguration is modeled as loss of capacity during the reconfiguration delay, where it is calculated based on a pre-defined function using the number of retunings as a parameter in [25], and not considered explicitly in [58].

Another related research problem arises in multihop WDM networks where each node is equipped with some small number of transmitters and receivers, each of which can communicate on one wavelength [21, 22, 23]. An assignment of independent transmit and receive wavelengths to each node defines the logical topology. Since the number of transmitters and receivers per node is limited, each source-destination pair cannot be assigned a unique wavelength and a message may have to hop through several intermediate nodes before finally reaching its destination. With the use of optically agile (slowly tunable) transceivers, the logical connectivity can be updated in response to changing traffic patterns. The problem is to find the network connectivity (wavelength assignment) and partition the flow of traffic among the links created (routing) such that the largest flow on any link is minimized. In [21], the wavelength assignment and routing problems are formulated jointly as a MILP. The authors also provide a heuristic method which decomposes the joint problem into two subproblems and iterates between them. The connectivity subproblem is formulated as an ILP and the routing subproblem is modeled as a multi-commodity flow problem. Same problem is also considered in [23], where minimum hop routing is assumed. The logical topology is incrementally adjusted towards a desired configuration using branch exchange algorithms. The reconfigurations are initiated at regular intervals and the branch exchange action which provides largest reduction in the maximum link load is applied. The problem is extended in [22] by associating a nonnegligible overhead with WDM reconfiguration in the sense that tuned transceivers cannot service backlogged data. For the joint solution of WDM reconfiguration and IP layer routing, several algorithms based on maximum weighted matchings are proposed and asymptotic throughput optimality of these algorithms are demonstrated.

A theoretically similar resource allocation problem is studied in the context of computing grid architectures in [41]. In that work, resources are the servers and competing jobs join separate queues based on their type. Available servers are grouped into clusters to serve different types of jobs. The servers are dynamically

switched between clusters in order to minimize the holding cost of jobs in the system. Switching of servers result in periods of time during which the corresponding servers remain idle and the objective is to minimize the total weighted holding cost of jobs. An MDP formulation is developed for the problem and optimum switching policies are obtained. A heuristic method is also proposed and shown to produce efficient resource utilization compared to static allocation. The ideas considered in [41] are also adapted to the DWA problem and used to develop alternative solutions for comparison with proposed methods in this work.

Chapter 4

Exact Solution of the DWA Problem

In this chapter, the DWA problem introduced in Chapter 3 is modeled as a continuous-time MDP with appropriate definitions of state representation, action space, state transition rates and cost function. Exact solution of the problem is obtained by converting the resulting model into an equivalent discrete-time process and solving it using the numerical technique of *value iteration.* Several cost function are also proposed and compared through simulations.

4.1 MDP Model

Using the network model introduced in Section 3.4.3, the DWA problem can well be represented with a continuous-time Markov chain whose state transition rates depend on the reconfiguration actions taken. The resulting process is called an MDP [61]. The formal definitions of state representation, action space, state transition rates and cost are given in the following subsections.

4.1.1 State Representation

The state of the network, $s \in S$, can be represented by the triplet $s = (\boldsymbol{f}, \boldsymbol{w}, k)$. $\boldsymbol{f} = [f_i]$ is the flow vector, where f_i is the number of flows at node i, $\boldsymbol{w} = [w_i]$ is the wavelength vector, where w_i is the number of wavelengths allocated to node i, and k indicates the node to which a wavelength is currently being switched. If no switching action is underway, k is 0. Valid states are $s = (\boldsymbol{f}, \boldsymbol{w}, k)$ such that $I_+(k) + \sum_i w_i = W$, where $I_+(k) = 1$ if $k > 0$, $I_+(k) = 0$ otherwise.

4.1.2 Action Space

A_s is the action space consisting of the valid actions that may be taken at state s. Since at most one switch at a time is allowed, if there is already a wavelength being switched, i.e., if $k > 0$, $A_s = \{a_0\}$, where a_0 corresponds to no-switching. Otherwise, A_s consists of a_0 and subset of actions a_{lm}, which correspond to switching one wavelength from node l to node m, such that node l has more than one wavelength allocated. That is,

$$A_s = \begin{cases} \{a_0\}, & \text{if } k > 0 \\ \{a_0\} \cup \{a_{lm} : w_l > 1, \quad l, m = 1, \ldots, N\}, & \text{if } k = 0. \end{cases} \tag{4.1}$$

4.1.3 State Transition Rates

Transition rates from state $s = (\boldsymbol{f}, \boldsymbol{w}, k)$ to state $s' = (\boldsymbol{f'}, \boldsymbol{w'}, k')$ under action a is denoted as $q_{ss'}(a)$. Transition rates when no switching is performed are given as

$$q_{ss'}(a_0) = \begin{cases} \lambda_i, & \text{if } f' = f + e_i \\ w_i \mu_i, & \text{if } f' = f - e_i \\ \sigma, & \text{if } k > 0 \text{ and } k' = 0 \text{ and } w' = w + e_k \\ 0, & \text{otherwise,} \end{cases}$$

where e_i denotes the unit vector which has 1 in position i and zeros elsewhere. When the action is to switch a wavelength from node l to node m (a_{lm}), there

is an instant transition from state $s = (\boldsymbol{f}, \boldsymbol{w}, k)$ to state $s' = (\boldsymbol{f'}, \boldsymbol{w'}, k')$, with $w' = w - e_l$ and $k' = m$.

4.1.4 Cost Function

The cost function is represented as $g(s, a)$, where s is the state and a is the action. It defines the cost per unit time, depending on the state of the system and/or the action taken. The objective of the MDP is to minimize the infinite horizon total discounted cost defined as:

$$\lim_{n\to\infty} \mathrm{E}\left\{\int_0^{t_n} e^{-\beta t} g\left(s\left(t\right), a\left(t\right)\right) \mathrm{d}t\right\},$$

where t_n is the occurrence time of the n^{th} state transition and β is the discount rate [62].

4.2 Uniformization of the MDP Model

In order to solve the continuous time MDP formulation, it is convenient to construct an equivalent discrete time process and use dynamic programming techniques. In the DWA problem, the control actions (wavelength switching) is applied at discrete times (flow arrival or departure instants), but the cost is continuously accumulated. Moreover, the time between successive control choices is variable and depends on the current state and the action taken, resulting in non-uniform transition rates. To develop the discrete time equivalent process, transition rates should be made uniform regardless of the state and the action. For this aim the technique of *uniformization* is used [62]. The basic idea of uniformization is to introduce fictitious transitions from a state to itself, so that the transitions that are slow on the average are speeded up with the added transitions.

The uniform transition rate, ν, should be greater than the maximum transition rate of the original process. Hence, for the continuous time MDP at hand, a

suitable choice may be

$$\nu = \sum_{i=1}^{N} \lambda_i + W\mu + \sigma$$

where $\mu = \max(\mu_i)$. Next, an equivalent discrete time Markov chain is constructed with the following transition probabilities:

$$p_{ss'}(a) = \begin{cases} q_{ss'}(a)/\nu, & \text{if } s' \neq s \\ 1 - q_s(a)/\nu, & \text{if } s' = s \end{cases}$$

where $q_s(a) = \sum_{s'} q_{ss'}(a)$.

The discount factor for the resulting discrete-time Markov chain is

$$\tilde{\beta} = \frac{\nu}{\beta + \nu}$$

The cost per stage is calculated as

$$\tilde{g}(s,a) = \frac{g(s,a)}{\beta + \nu}$$

Then, Bellman's equation takes the form

$$J(s) = \min_{a \in A_s} \left[\tilde{g}(s,a) + \tilde{\beta} \sum_{s'} p_{ss'}(a) J(s') \right] \tag{4.2}$$

where $J(s)$ is the cost associated with state s for a given policy [62].

4.3 Solution of the MDP Model

The solution of the set of linear equations in (4.2) results in the value of each state and the optimum switching policy, which is the action to be taken at each state. The number of flows at each node is a process described by the Markov chain depicted in Figure 4.1, where the service rate, μ, depends on the number of wavelengths allocated to the node. Since this chain is infinite, the size of the state space, S, and therefore the number of equations in (4.2) is infinite. In order to use numerical solution techniques, the number of equations should be made

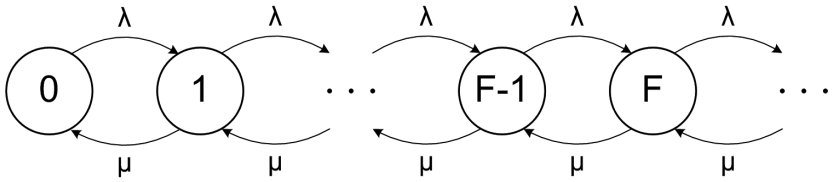


Figure 4.1: Infinite Markov chain.

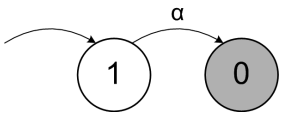


Figure 4.2: Exponential distribution.

finite. This can be accomplished by truncating the number of flows at each node at F. A simple truncation may be inadequate if the probability of states beyond F is not negligible. For this reason, it may be a better idea to match the first moment of the sojourn time in the truncation process.

For the infinite Markov chain with uniform transition rates, sojourn time at the set of states $\geq F$, for any value of F, is exactly same as the busy period in an M/M/1-PS queue. The first moment (i.e., mean) of this distribution is

$$m_1 = E[T] = \frac{1}{(1-\rho)}\frac{1}{\mu},$$

where $\rho = \lambda/\mu$.

The first moment of the sojourn time distribution can be matched using a simple exponential distribution (Figure 4.2) with rate $\alpha = 1/m_1$ [63]. The resulting chain is shown in Figure 4.3, where the state $F+$ corresponds to the set of states with number of flows equal to or greater than F. After truncation, the resulting finite-state discrete-time MDP is solved using the method of *value iteration* [64].

4.4 Cost Functions

Cost function is the key component of any optimization problem and it should be designed in accordance with the objectives of the optimization. For the DWA

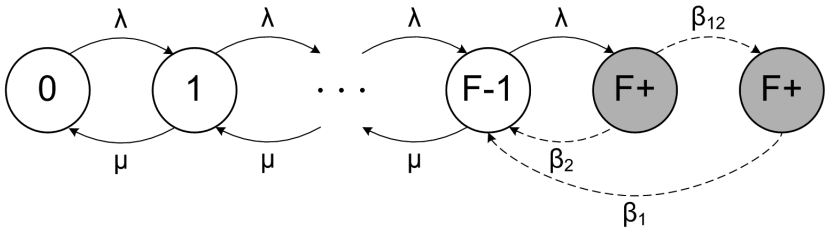


Figure 4.3: Truncated Markov chain with first moment matched.

problem at hand, the primary goal is to maximize the throughput which is equivalent to minimizing flow completion times. Meanwhile, it is also desirable that each flow uses a fair share of the capacity available. Following cost functions are considered in this thesis.

1. **Flow Sum (FS)**

   As discussed in Section 3.4.1.1, in resource allocation problems it is usually assumed that each unfinished job in the system incurs a holding cost with a specific rate. Without loss of generality this rate can be assumed to be unitary for each flow. Then, the holding cost for each time unit corresponds to the total number of flows in the network and can be stated as

   $$g(s,a) = g_{FS}(s) = \sum_i f_i. \tag{4.3}$$

   FS is also used in the optimization problem in [41]. As discussed in Section 3.4.1.1, minimum holding cost is achieved when the average flow throughput is maximized. Hence, the optimum policy minimizing FS results in the minimum holding cost and maximum average flow throughput.

2. **Normalized Flow Sum (NFS)**

   This cost function is derived from the heuristic method of [24]. Although not explicitly stated in that paper, this heuristic method can be seen as an approximation to the optimum policy obtained with the cost function

   $$g(s,a) = g_{NFS}(s) = \sum_i \frac{f_i}{w_i}.$$

   The motivation behind NFS is to balance the load between wavelength channels to achieve high throughput and fairness.

3. **Normalized Squared Flow Sum (NSFS)**

   In this work, we propose the following cost function

$$g(s,a) = g_{NSFS}(s) = \sum_i \frac{f_i^2}{w_i}.$$

   The basic idea behind NSFS is to minimize both the flow completion times and load imbalance between wavelength channels in order to obtain better results in terms of throughput and fairness.

These cost functions may be compared based on the following properties, which can be considered useful in order to achieve the objectives of throughput and fairness.

P1. *Cost of a node should be an increasing function of number of flows at the node.* This property is based on the idea that the throughput can be maximized by minimizing the duration of flows at each node. All of the above cost functions satisfy this property.

P2. *Cost of a node should be a decreasing function of number of wavelengths assigned to the node.* It is clear that, increasing the service rate also increases the throughput. Moreover, this property is useful to account for the costs associated with the unavailability of the reconfigured wavelength during the reconfiguration period. It is observed that this property holds for the cost functions NFS and NSFS.

P3. *Total cost should be minimum when the load is balanced among wavelength channels.* A fair service is achieved when the number of wavelengths at each node is proportional to the number of flows at the corresponding node. Hence, it may be desirable that the cost function attains the minimum value at this point, i.e., when w_i is proportional to f_i. This property is satisfied only by the cost function NSFS as shown in the following lemma.

   Lemma 2. *The function,* $g(\boldsymbol{f}, \boldsymbol{w}) = \sum_i f_i^2/w_i$*, is convex cup, and it is minimized when* $\boldsymbol{w}$ *is proportional to* $\boldsymbol{f}$*.*

Proof. Let R be the region consisting of $\boldsymbol{w}$ vectors defined by

$$\sum_{i=1}^{N} w_i = W$$

For any vector $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ in R, the vector $\theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}$ is in R for $0 \leq \theta \leq 1$, because

$$\sum_{i=1}^{N} \left(\theta\alpha_i + (1-\theta)\,\beta_i\right) = \theta W + (1-\theta)W = W$$

So, R is a convex region. For all $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ in R and $0 \leq \theta \leq 1$,

$$\begin{aligned} \theta g(\boldsymbol{f}, \boldsymbol{\alpha}) + (1-\theta) g(\boldsymbol{f}, \boldsymbol{\beta}) - g(\boldsymbol{f}, \theta\boldsymbol{\alpha} + (1-\theta)\boldsymbol{\beta}) &= \\ \sum_{i=1}^{N} f_i^2 \frac{(\alpha_i - \beta_i)^2}{\alpha_i \beta_i (\theta\alpha_i + (1-\theta)\beta_i)} &\geq 0 \end{aligned}$$

Hence, g is convex cup ($\cup$) over R and therefore it has a minima which can be found using the method of Lagrange Multipliers:

$$\frac{\partial}{\partial w_i}\left(\sum_{i=1}^{N} \frac{f_i^2}{w_i} + L\left(\sum_{i=1}^{N} w_i - W\right)\right) = -\frac{f_i^2}{w_i^2} + L = 0$$

$$\implies \frac{f_i^2}{w_i^2} = L$$

$$\implies \frac{f_i}{w_i} = \sqrt{L}$$

□

In the following section these cost functions are used in the MDP formulation to obtain optimum reconfiguration policies and the performance of these policies are compared based on the metrics defined in Section 3.4.1.

4.5 Comparison of Cost Functions

The three cost functions are compared on a 3-node network scenario shown in Figure 4.4. In this network, there are 7 wavelength channels. Flow arrival rates

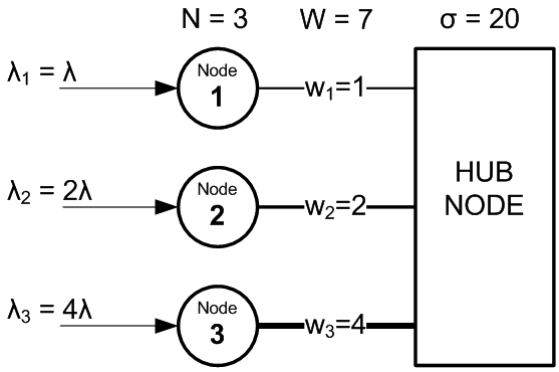


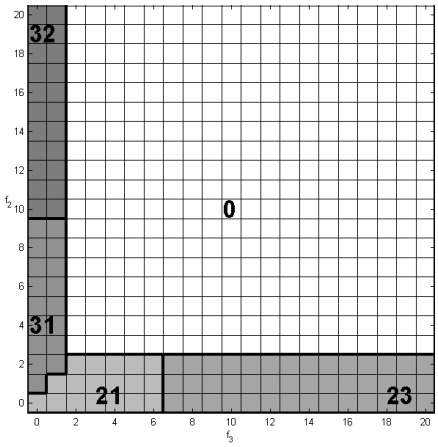
Figure 4.4: 3-node test network.

are λ, 2λ, and 4λ flows/s to nodes 1, 2, and 3, respectively. The bandwidth of a single channel is 10 Gpbs and the average flow size is 1250 MB. Hence, the service rate of a flow by a single channel, μ_i, is 1 flows/s, for all nodes i. Average reconfiguration delay, $1/\sigma$, is 50 ms.

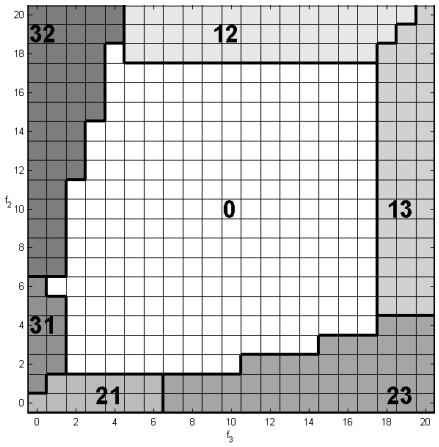
Figures 4.5a, 4.5b and 4.5c show parts of the optimum policies (corresponding to states with $w = [3, 2, 2]$ and $f = [15, f_2, f_3]$) obtained using the cost functions considered, for $\lambda = 0.7$ and $F = 20$. The x-axis corresponds to the number of flows at node 3 and the y-axis is the number of flows at node 2. Each cell in the matrices represents a single state and the value of the cell is the optimum action to be taken at that state. The switching actions, a_{ij} are labeled on the figures as ij meaning that a switch from node i to node j is to be performed. No switching decisions, a_0, are labeled as 0.

FS aims to minimize the total duration of flows in the network and it does not consider load balancing at all. Consistent with this objective, it prefers to switch a wavelength from a node when the number of flows at that node is very low, as can be observed from Figure 4.5a. Since the number of flows at node 1 is large for all states shown, the policy does not make switches from node 1.

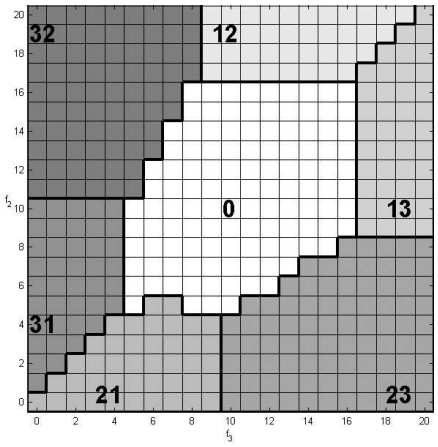
The policy obtained using NFS is shown in Figure 4.5b. It is observed that this policy makes more switches compared to FS policy. In addition to the actions taken when the number of flows at a node gets small, this policy also makes switches to achieve load balancing between wavelengths. This is evident from the



(a) Optimum policy for FS



(b) Optimum policy for NFS



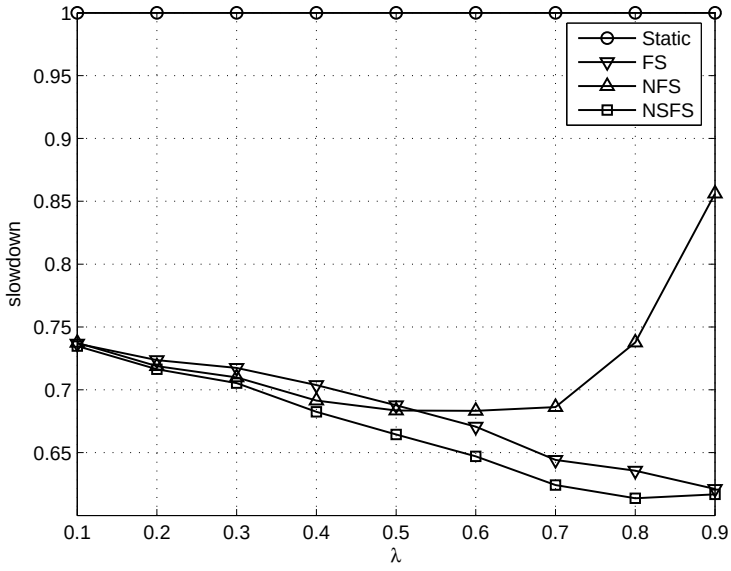
(c) Optimum policy for NSFS

Figure 4.5: Optimum switching policies for the 3-node test network, for states with $w = [3, 2, 2]$ and $f = [15, f_2, f_3]$.

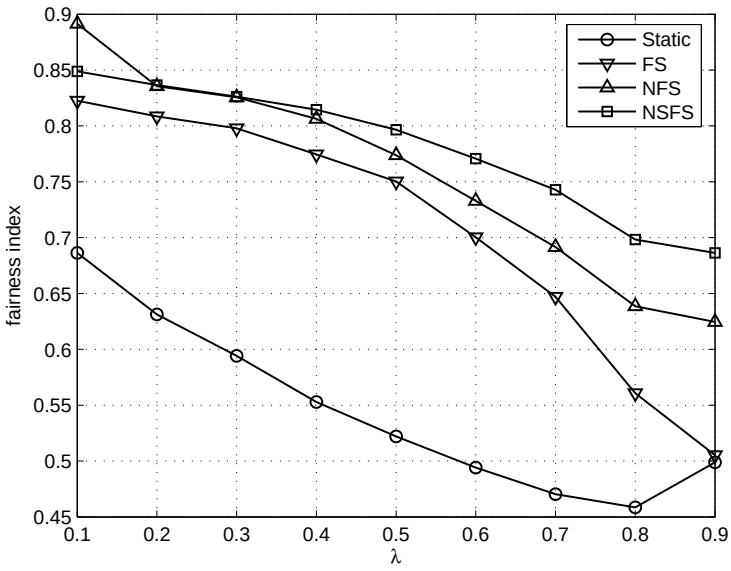
fact that, for high number of flows at nodes 2 or 3, a wavelength is switched from node 1.

Figure 4.5c plots the policy obtained with NSFS. This policy has a similar structure with the NFS policy, but the area corresponding to no action (a_0) is smaller. So, it may be concluded that this policy makes more switches in order to balance the load at each node.

In order to evaluate the performance of each policy, simulations are performed where λ is changed from 0.1 to 0.9. Each simulation is repeated 10 times and

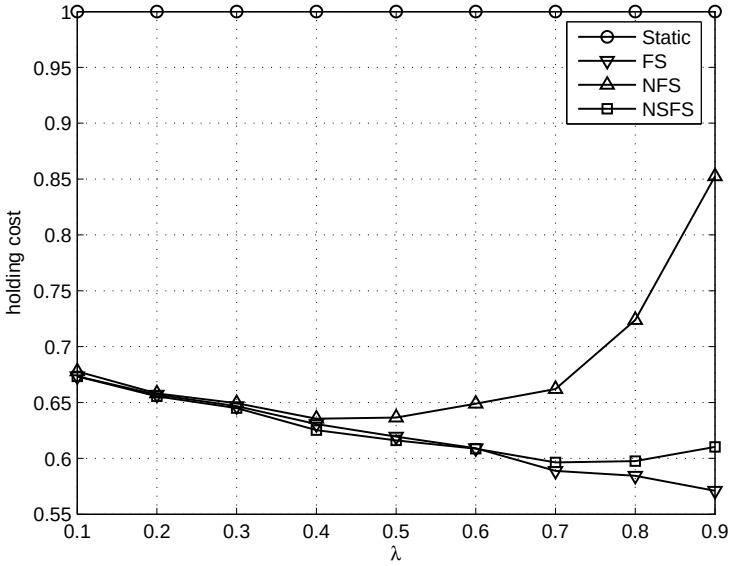


(a) Slowdown

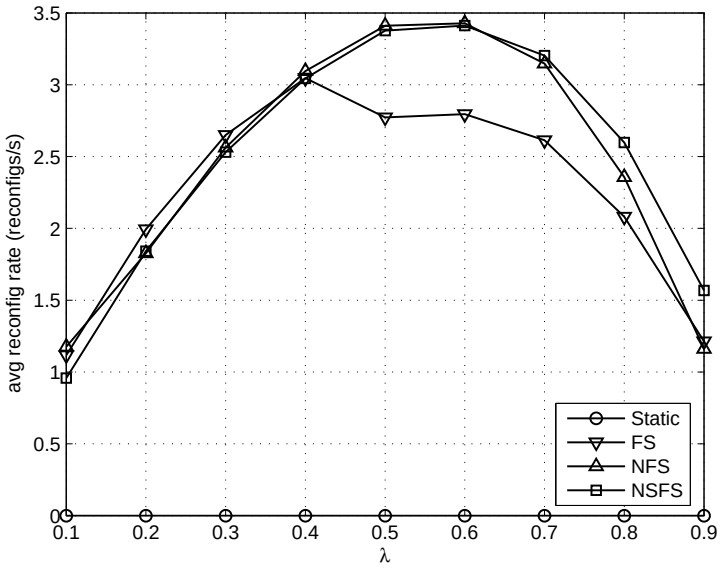


(b) Fairness

Figure 4.6: Performance of cost functions as a function of network load.



(c) Holding Cost



(d) Average Switching Rate

Figure 4.6 (continued): Performance of cost functions as a function of network load.

the average values are plotted. For comparison, we also use the static wavelength allocation, where the channels are assigned based on average traffic demands and they are not reconfigured. For this scenario, the static allocation corresponds to allocating 1, 2, and 4 wavelengths to nodes 1, 2, and 3, respectively. Figure 4.6a plots the slowdown obtained using each of the switching policies normalized with respect to the slowdown experienced under the static policy. As a first observation, it is seen that the dynamic policies yield significantly better slowdown performance than the static policy. Among the dynamic policies, NSFS achieves the minimum slowdown for all values of network load. The results obtained with FS are close to NSFS. On the other hand, the slowdown obtained with the cost function NFS gets worse as the network load increases.

The performances of policies in terms of fairness as a function of the flow arrival rate are compared in Figure 4.6b. Static policy has a clear disadvantage in terms of fairness. All of the dynamic policies have better fairness at low load levels but as the load increases the fairness begins to drop. Among the dynamic policies, worst performance belongs to FS. This is expected since FS does not consider load balancing. With this policy, fairness drops sharply at high loads to the level obtained by the static policy. NFS is better than FS, but NSFS shows the best performance except at very low load levels.

Figure 4.6c plots the holding cost obtained with each policy normalized to the holding cost experienced under the static policy. This graph is similar to the slowdown results. For low load levels all of the dynamic policies achieves 30%-35% lower holding costs compared to the static policy. The gains obtained with FS and NSFS increase further with the increasing load while the performance of NFS degrades sharply. It is observed that although NSFS policy is better than FS in terms of slowdown, the difference is negligible in terms of the holding cost. In fact, FS policy optimizes the holding cost and the results suggest that NSFS policy reduces the slowdown and increases fairness without sacrificing the holding cost objective.

Average rate of reconfigurations (reconfigs per second) performed by each of the policies is depicted in Figure 4.6d. The curves corresponding to each policy

has a similar pattern. Reconfiguration rate increases with increasing load up to 0.6, and then begin to decrease with the increasing load. At moderate loads FS performs the minimum number of switches among all policies. This result is related to the fact that at moderate and high load levels, the probability of having small number of flows at any node decreases and FS policy does not make switches unless the number of flows at a node is very low.

In summary, it can be stated that the NSFS policy has important advantages as a DWA method. It attains minimum slowdown and maximum fairness nearly for all levels of network load.

Chapter 5

Heuristic Methods for the DWA Problem

As usual with most of the optimization problems, dynamic programming solution of the MDP, presented in Chapter 4, suffers from *curse of dimensionality*. The size of the state space for the MDP formulation is $F^N \times W_{max}^N \times (N+1)$, where N is the number of nodes, F is the truncation level for number of flows and W_{max} is the maximum number of wavelengths possible at each node, respectively. The size of the state space grows exponentially with N and the problem is solvable only for small networks. For a real-life IP/WDM network with around 10 nodes and 10-100 wavelengths, it is practically impossible to obtain a solution using the MDP approach.

In this chapter, three heuristic methods that can be used for the solution of the DWA problem are introduced. The first two methods (HM1 and HM2) are adapted from the heuristics found in the literature. The third one (HM3) is a new algorithm proposed in this study. An efficient implementation method for HM3 is also discussed.

5.1 Heuristic Method 1 (HM1)

This heuristic is inspired by the method devised in [41]. Although the context is different in that work, the underlying problem is similar to the DWA problem. HM1 makes switching actions if the action would help to balance the holding costs, taking into account the reconfiguration overheads. HM1 can be seen as an approximation to the optimum policy obtained using cost function FS.

Let $s^* = (\boldsymbol{f^*}, \boldsymbol{w^*}, k^*)$ denote the state of the network prior to any switching action and A_{s^*} be the valid set of actions at state s^*, as defined in (4.1). At each decision instant the following rule is applied:

- Calculate the following for each action $a_{ij} \in A_{s^*}$

$$v_{ij} = \left(f_j^* + \frac{1}{\sigma} \left(\lambda_j - \mu_j w_j^* \right) \right) - K \left(f_i^* + \frac{1}{\sigma} \left(\lambda_i - \mu_i (w_i^* - 1) \right) \right) \tag{5.1}$$

  where K is recommended to be 5 in [41].

- Take the action which yields the maximum v_{ij}, if it is strictly greater than 0.

Note that the departure rate term in [41] is appropriately modified and the holding costs of flows at each node is taken as 1, to adapt the heuristic to the problem at hand.

This method tries to estimate the effects of a switch. The first term in (5.1) is an approximation for the number of flows at node j after the reconfiguration delay. Similarly, the second term is the expected number of flows at node i after the reconfiguration is completed, scaled by a factor in order to discourage too many switches. Thus, the value of an action is related to the expected decrease in holding costs.

5.2 Heuristic Method 2 (HM2)

This heuristic is proposed in [24]. At each decision epoch, action $a_{ij} \in A_{s^*}$ with $i = \arg\min_x\{f_x^*/w_x^*\}$ and $j = \arg\max_x\{f_x^*/w_x^*\}$, is performed if the following inequality holds

$$\frac{f_j^*}{w_j^*+1} + \frac{f_i^*}{w_i^*-1} < \frac{f_j^*}{w_j^*} + \frac{f_i^*}{w_i^*}$$

where $s^* = (\boldsymbol{f^*}, \boldsymbol{w^*}, k^*)$ is the state of the network and A_{s^*} is the valid set of actions at state s^*

The basic idea behind HM2 is to keep all wavelengths in the system evenly loaded. So, a switch will be performed if it is going to improve the load balance in the system. The reconfiguration costs are not taken into account and the flow arrival and departure rates are not considered. HM2 may be thought as a first order approximation to the optimum policy obtained using cost function NFS.

5.3 Heuristic Method 3 (HM3)

It is shown in Chapter 4 that NSFS has desirable properties and among alternative cost functions it achieves the best performance in terms of both slowdown and fairness at all levels of network load. Therefore, it is intuitive to consider a heuristic approach which aims to minimize the cost function NSFS for the solution of the DWA problem. Based on this reasoning a new heuristic method, HM3, is developed and outlined below.

Denoting the state of the network at the decision instant as $s^* = (\boldsymbol{f^*}, \boldsymbol{w^*}, k^*)$ and the valid set of actions at state s^* as A_{s^*}, HM3 applies the following rule to determine the switching action:

- For each action $a_{ij} \in A_{s^*}$ calculate v_{ij} as

$$v_{ij} = 1 - F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*),$$

where λ_i and and λ_j are the arrival rates of flows at node i and node j, respectively. μ_i and μ_j are the service rate of flows by a single wavelength channel at node i and node j, respectively. $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ is the probability that in a duration of τ after applying a_{ij}, a state will be reached for which the prospective wavelength allocation is worse than the original allocation in terms of NSFS cost. Therefore, v_{ij} represents the probability that at all of the states which will be visited during the reconfiguration period, the prospective wavelength allocation will have a smaller NSFS cost then the original wavelength allocation.

- Apply the action with maximum v_{ij} (v_{max}), if $v_{max} > V_{thr}$, where $0 < V_{thr} < 1$ is the switching threshold, used in order to eliminate unnecessary reconfigurations. Unless otherwise stated V_{thr} is set to 0.85 in this work. The sensitivity of performance with respect to the V_{thr} value is evaluated in Section 6.3.

In the rest of this section $F_\tau(*)$ is used as a shorthand notation for $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$. The basic reasoning behind HM3 is discussed in the following section. In Section 5.3.2 the method used to calculate the function $F_\tau(*)$ is explained. And finally an efficient implementation of $F_\tau(*)$ is considered in Section 5.3.3.

5.3.1 Geometric Interpretation of HM3

As long as the NSFS cost is considered, the action a_{ij} affects the costs incurred only at node i and node j. Therefore, it is sufficient to consider these nodes to evaluate the potential consequences of the switching action a_{ij}. Number of flows at the nodes evolve according to independent processes which may be represented by separate M/M/1-PS queueing models and the joint process can be described by a two dimensional Markov chain. The states of the chain correspond to the number of flows at node i and node j, i.e., (f_i, f_j).

A sample Markov chain is shown in Figure 5.1 for the case where the total

number of wavelengths allocated to the node i and node j is 4. Note that the chain is infinite in both f_i and f_j directions and only a part of it is shown in the figure. Each circle corresponds to a state with f_j and f_i values given x- and y-axis, respectively. Arrivals to node i (node j) are exponential with rate λ_i (λ_j) and correspond to transitions in the north (east) direction. Similarly, departures from node i (node j) are exponential with rate $w_i\mu_i$ ($w_j\mu_j$) and result in transitions to the states in south (west) direction.

Each possible wavelength allocation (w_i, w_j) corresponds to a line, $L_{w_iw_j}$, with slope w_i/w_j. For the states on line $L_{w_iw_j}$ a perfect load balance is achieved between the nodes i and j with the wavelength allocation (w_i, w_j). As the state moves away from the line corresponding to the actual wavelength allocation, load imbalance increases and another wavelength allocation may become more beneficial in terms of NSFS cost. Hence, the state space can be divided into regions in each of which a different wavelength allocation is preferable. Each region where (w_i, w_j) results in minimum cost are labeled as R_{w_i,w_j} in the figure. For instance if $f_i = 5$ and $f_j = 10$ then it is observed that allocating 2 wavelengths to each node results in the lowest cost.

For a state (f_i, f_j), the wavelength allocation $(w_i - 1, w_j + 1)$ results in less NSFS cost than the wavelength allocation (w_i, w_j) if

$$\frac{f_i^2}{w_i - 1} + \frac{f_j^2}{w_j + 1} < \frac{f_i^2}{w_i} + \frac{f_j^2}{w_j},$$

which may be rewritten as

$$\frac{f_i}{f_j} < m = \sqrt{\frac{w_i(w_i - 1)}{w_j(w_j + 1)}}. \tag{5.2}$$

(5.2) corresponds to the lines separating regions in the Figure 5.1.

Assuming that, at the decision instant $f_i = f_i^*$, $f_j = f_j^*$ and $w_i = w_i^*$ and $w_j = w_j^*$, the wavelength allocation $w_i = w_i^* - 1$ and $w_j = w_j^* + 1$ results in less NSFS cost if $f_i^* < m^* f_j^*$, where m^* is calculated using the wavelength distribution at the decision instant

$$m^* = \sqrt{\frac{w_i^*(w_i^* - 1)}{w_j^*(w_j^* + 1)}}. \tag{5.3}$$

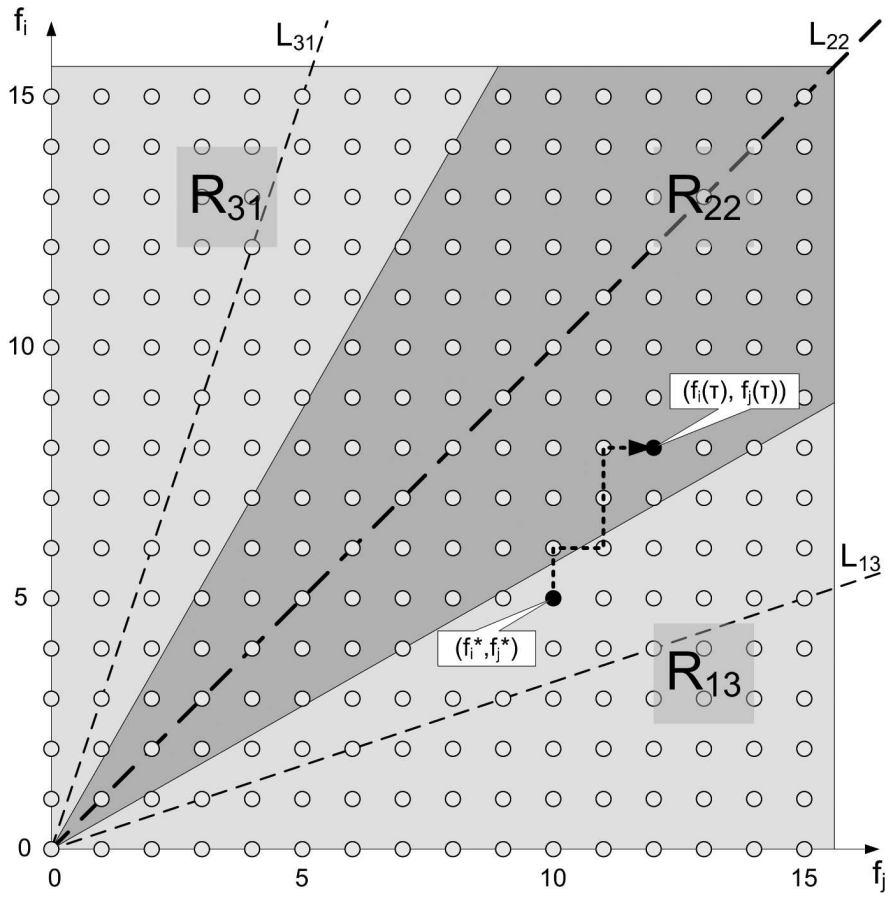


Figure 5.1: Geometric interpretation of HM3.

In this case, if the wavelength switching actions are instantaneous, i.e. $\tau = 0$, then a_{ij} should be applied since it decreases the cost.

However, in the DWA problem there is a non-zero reconfiguration delay associated with the reconfiguration actions. The time interval of length τ between the initiation of the switching action and reaching the final wavelength allocation is called the *reconfiguration period*. As soon as the action a_{ij} is applied wavelength allocation changes to $(w_i^* - 1, w_j^*)$ and stays so in the reconfiguration period only after which the switching is completed and the wavelength allocation finally changes to $(w_i^* - 1, w_j^* + 1)$. The existence of the reconfiguration period introduces several issues.

First, during the reconfiguration period the capacity of a whole wavelength channel remains idle and a cost is incurred since the total service rate of the

network decreases. Therefore, the action should be applied only if the long term rewards outweighs this cost.

Second, since at most one wavelength is allowed to be in the switching state at any time, application of a_{ij} prevents the application of other possibly useful actions during the reconfiguration period. Therefore most beneficial action should be applied first.

Third, during the reconfiguration period, f_i and f_j evolve in time $0 \leq u \leq \tau$ according to

$$f_i(u) = f_i(0) + a_i(u) - d_i(u) \tag{5.4}$$

$$f_j(u) = f_j(0) + a_j(u) - d_j(u) \tag{5.5}$$

where u is the time passed after the application of the action, i.e., $u = 0$ at the decision instant, and $f_i(0) = f_i^*$ and $f_j(0) = f_j^*$. $a_i(u)$ ($a_j(u)$) and $d_i(u)$ ($d_j(u)$) are the number of arrivals and departures at node i (node j) in the reconfiguration period up to time u, respectively. Even if the wavelength allocation (w_i^*-1, w_j^*+1) decreases the NSFS cost for the state (f_i^*, f_j^*) at the decision instant, for some of the states $(f_i(u), f_j(u))$ visited during the reconfiguration interval the wavelength allocation (w_i^*, w_j^*) may turn out to be more efficient. Moreover, the state reached at the end of the reconfiguration period, $(f_i(\tau), f_j(\tau))$, may require the reverse action a_{ji} just to undo the last switching action, incurring additional cost to return back to the previous wavelength distribution.

To demonstrate this phenomena a sample path of state evolution is plotted on Figure 5.1, where the wavelength allocation (w_i^*, w_j^*)=(2,2) and state $(f_i^*, f_j^*) = (5, 10)$ at the decision instant. Assume that a wavelength is switched from node i to node j since $(f_i, f_j) \in R_{1,3}$. However, during the reconfiguration period, the state moves into R_{22} and at the end of the reconfiguration period the final state turns out to be (8,12) for which the original wavelength allocation (2,2) is indeed more efficient. As a result, the action a_{ij} increases the NSFS cost in addition to the incurred capacity loss. Moreover, it requires another action a_{ji} to turn back to the wavelength allocation of (2,2). It also prevents the application of possible useful actions during the reconfiguration period.

To capture all these three issues, HM3 considers the possible state trajectories that may be followed during the reconfiguration period. It attaches a value, $v_{ij} = 1 - F_\tau(*)$ to action a_{ij}, which is equal to the probability that the resulting wavelength allocation decreases the cost for all the states visited during the reconfiguration period. Then, HM3 applies the action with maximum v_{ij} value, if this value is greater than the switching threshold, V_{thr}. For the case illustrated in Figure 5.1, v_{ij} corresponds to the probability that the sample path stays in region R_{13} during the reconfiguration period.

The basic idea is that, as the distance of (f_i^*, f_j^*) to $L_{w_i^* w_j^*}$ gets larger, the load imbalance between node i and node j and hence the potential benefits of reconfiguration increases. The value of v_{ij} also increases as (f_i^*, f_j^*) moves away from $L_{w_i^* w_j^*}$ and the action with maximum v_{ij} may be thought to improve the cost most. Moreover, a larger value of v_{ij} may also indicate a greater probability that the long term benefits of switching overweighs the reconfiguration costs and a switching threshold, V_{thr}, may be used to eliminate unnecessary actions. Finally the probability that the prospective wavelength allocation will be efficient during and at the end of the reconfiguration period also increases with v_{ij}.

5.3.2 Calculation of $F_\tau(*)$

$F_\tau(*)$ is defined as the probability that during the reconfiguration period following the action a_{ij}, (f_i, f_j) changes so that the NSFS cost with the prospective wavelength allocation $(w_i^* - 1, w_j^* + 1)$ is greater than the NSFS cost with the original wavelength allocation (w_i^*, w_j^*), at least for some point in time. Using (5.3), this probability can be written as

$$F_\tau(*) = \Pr\left(f_i(u) > m^* f_j(u) \text{ for some } u \in [0, \tau) \mid f_i(0) = f_i^*, f_j(0) = f_j^*\right) \quad (5.6)$$

where $f_i(u)$ and $f_j(u)$ are defined in (5.4) and (5.5), respectively. So, $F_\tau(*)$ can be expressed as

$$F_\tau(*) = \Pr\left(T_{cD} < \tau \mid c = (f_i^*, f_j^*), D = ((f_i, f_j) : f_i > m^* f_j)\right) \quad (5.7)$$

where T_{cD} is the time that starting from c, the first passage to a state in D occurs in the two-dimensional birth-death process with arrival rate λ_i, (λ_j), and service

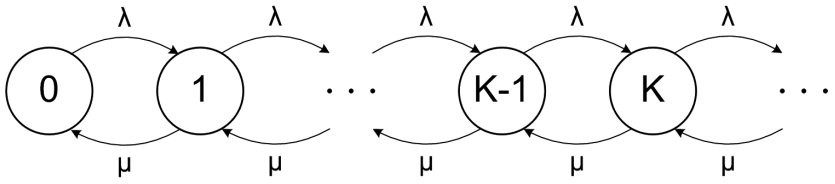


Figure 5.2: Infinite Markov chain.

rate $(w_i - 1)\mu_i$ $(w_j\mu_j)$ at node i (node j). τ is the reconfiguration delay, which has an exponential distribution with rate σ. If $f_i^* > m^* f_j^*$ then it follows directly that $F_\tau(*) = 1$, since the condition in (5.6) holds for $u = 0$ with certainty.

Probability density function of T_{cD} can be calculated using numerical techniques which are applicable to finite Markov chains. In order to have a finite two-dimensional chain, the birth death processes at node i and node j can be truncated at levels F_i $(> f_i^*)$ and F_j $(> f_j^*)$, respectively. For this aim, the method described in Note 5.3.1 is used with $K = F_i$, $\lambda = \lambda_i$, $\mu = (w_i - 1)\mu_i$ and $K = F_j$, $\lambda = \lambda_j$, $\mu = w_j\mu_j$ for node i and node j, respectively.

Note 5.3.1 (Truncation of a Markov Chain with Three Moment Matching)**.** *For an infinite Markov chain with uniform transition rates (Figure 5.2), sojourn time, T, at states $s \geq K$ for any K is exactly the same as the busy period in an M/M/1-PS queue. First three moments of T are:*

$$m_1 = E[T] = \frac{1}{(1-\rho)}\frac{1}{\mu}$$

$$m_2 = E[T^2] = \frac{2}{(1-\rho)^3}\frac{1}{\mu^2}$$

$$m_3 = E[T^3] = \frac{6(1+\rho)}{(1-\rho)^5}\frac{1}{\mu^3}$$

where $\rho = \lambda/\mu$.

In order to match the first three moments of the sojourn time, two-phase Coxian$^+$PH distribution, shown in Figure 5.3, can be used [63].

The parameters of Coxian$^+$PH distribution are

$$\beta_1 = (1 - p_x)\lambda_{x1} \qquad \beta_{12} = p_x\lambda_{x1} \qquad \beta_2 = \lambda_{x2}$$

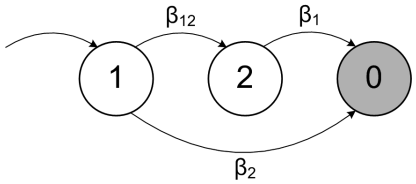


Figure 5.3: $Coxian^{+}PH$ distribution.

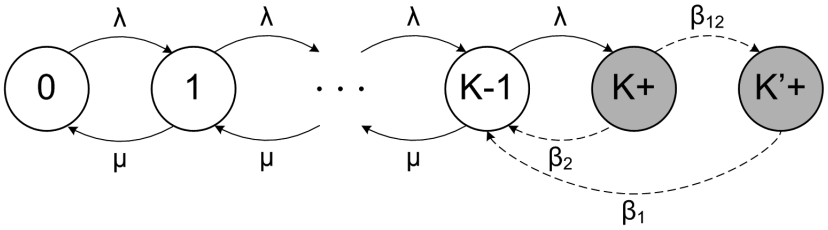


Figure 5.4: Truncated Markov chain with first three moments matched.

$$\lambda_{x1} = \frac{u + \sqrt{u^2 - 4v}}{2\mu_1} \qquad \lambda_{x2} = \frac{u - \sqrt{u^2 - 4v}}{2\mu_1}$$

$$p_x = \frac{\lambda_{x2}(\lambda_{x1}\mu_1) - 1}{\lambda_{x1}}$$

$$u = \frac{6 - 2m_3}{3m_2 - 2m_3} \qquad v = \frac{12 - 6m_2}{m_2(3m_2 - 2m_3)}$$

The resulting truncated chain is shown in Figure 5.4.

After the truncation, Markov chains associated with node i and node j have $F_i + 2$ and $F_j + 2$ states, respectively. The generator matrix, Q, for the finite two-dimensional chain corresponding to the joint process is constructed using these truncated chains. With c and D defined in (5.7), the Laplace Transform of the first passage time probability density function is calculated using the method described in Note 5.3.2.

Note 5.3.2 (First Passage Time Probability Distribution)**.** *For a finite, irreducible, continuous time Markov chain (CTMC) with n states and generator matrix Q, the first passage time from a source state c into a non-empty set of target states D is defined as*

$$T_{cD}(t) = inf\{u > 0 : X(t + u) \in D \mid X(t) = c\}$$

where $X(t)$ denotes the state of CTMC at time $t \geq 0$ [65].

When the CTMC is stationary and time-homogeneous, T_{cD} is independent of t:

$$T_{cD} = inf\{u > 0 : X(u) \in D \mid X(0) = c\}$$

Let $f_{cD}(t)$ be the probability density function of T_{cD}, then

$$\Pr(a < T_{cD} < b) = \int_a^b f_{cD}(t)\mathrm{d}t \qquad 0 \leq a < b$$

Using a first step analysis, the Laplace transform of f_{cD} can be written as

$$L_{cD}(s) = \sum_{k \notin D} p_{ck} \left(\frac{-q_{cc}}{s - q_{cc}} \right) L_{kD}(s) + \sum_{k \in D} p_{ck} \left(\frac{-q_{cc}}{s - q_{cc}} \right)$$

The first term denotes the event that the system first transits to a non-target state k then to a target state in D. The second term is for the case where the system transits from state c directly to a state in D. Using the relation $p_{ck} = -q_{ck}/q_{cc}$, this expression can be rewritten as,

$$(s - q_{cc})L_{cD}(s) = \sum_{k \notin D} q_{ck} L_{kD}(s) + \sum_{k \in D} q_{ck} \tag{5.8}$$

The set of equations can also be expressed in matrix-vector form. For example, when $D = \{1\}$,

$$\begin{bmatrix} s - q_{11} & -q_{12} & \cdots & -q_{1n} \\ 0 & s - q_{22} & \cdots & -q_{2n} \\ 0 & -q_{32} & \cdots & -q_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -q_{2n} & \cdots & s - q_{nn} \end{bmatrix} \begin{bmatrix} L_{1D}(s) \\ L_{2D}(s) \\ L_{3D}(s) \\ \vdots \\ L_{nD}(s) \end{bmatrix} = \begin{bmatrix} 0 \\ q_{21} \\ q_{31} \\ \vdots \\ q_{n1} \end{bmatrix}$$

The value of $L_{cD}(s)$ can be obtained by solving this set of n linear equations. Note that the solution of (5.8) provides $L_{cD}(s)$ values for each state c of the Markov chain.

To obtain $\Pr(T_{cD} < \tau)$, one way is to calculate the following integral numerically

$$\begin{aligned}\Pr(T_{cD} \leq \tau) &= \int_0^\infty \Pr(T_{cD} \leq x) \Pr(\tau = x) \mathrm{d}x \\ &= \int_0^\infty \left(\int_0^x f_{cD}(t) \mathrm{d}t \right) \sigma e^{-x\sigma} \mathrm{d}x\end{aligned}$$

where $f_{cD}(t)$ values for the required values of t can be obtained using one of several methods for numerical transform inversion, such as Euler and Post-Widder algorithms [66]. These methods necessitate the calculation of $L_{cD}(s)$ at a number of different s values, depending on the desired accuracy. Fortunately, the Laplace transform L_{cD} has also a direct probabilistic interpretation as shown in the following lemma.

Lemma 3. *$L_{cD}(\sigma) = \Pr(T_{cD} < \tau)$ where τ is an exponential random variable with rate σ.*

Proof.

$$\begin{aligned}\Pr(T_{cD} \leq \tau) &= \int_0^\infty \Pr(T_{cD} \leq t) \sigma e^{-t\sigma} \mathrm{d}t \\ &= \int_0^\infty \int_0^t f_{cD}(s) \mathrm{d}s \sigma e^{-t\sigma} \mathrm{d}t \\ &= \int_0^\infty f_{cD}(s) \int_s^\infty \sigma e^{-t\sigma} \mathrm{d}t \mathrm{d}s \\ &= \int_0^\infty f_{cD}(s) e^{-s\sigma} \mathrm{d}s \\ &= L_{cD}(\sigma)\end{aligned}$$

□

Using Lemma 3, the desired probability and hence $F_\tau(*)$ can simply be calculated as

$$F_\tau(*) = \Pr(T_{cD} < \tau) = L_{cD}(\sigma).$$

5.3.3 Efficient Implementation of $F_\tau(*)$

HM3 requires the calculation of $F_\tau(*)$ for each valid action $a_{ij} \in A_{s^*}$ at each decision instant according to the method described in Section 5.3.2. As will be explained shortly, it is also possible to estimate the values of $F_\tau(*)$ for all f_i^* and f_j^* based on the values calculated for $0 \leq f_i^* \leq F_i$ and $0 \leq f_j^* \leq F_j$ for some F_i and F_j. Hence, values of $F_\tau(*)$ for a representative set of states can be calculated off-line and these data can be utilized at each decision instant to obtain the required values of F_τ. In the rest of this section, the validity of this approach is discussed and related algorithms are presented.

The Markov chain used for the calculation of $F_\tau(*)$ is shown in Figure 5.5. L_T is the threshold line separating regions R_{w_i,w_j} and R_{w_i-1,w_j+1} and has a slope of m^* as defined in (5.3). Note that D corresponds the set of states above L_T. If $c = (f_i^*, f_j^*)$ is above L_T then by definition $F_\tau(*) = 0$. Hence, the values of $F_\tau(*)$ are to be determined for states $c = (f_i^*, f_j^*)$ below L_T. (f_i, f_j) evolve during the reconfiguration period according to (5.4) and (5.5). The flow departure (arrival) rates are $(w_i^* - 1)\mu_i$ (λ_i) and $w_j^*\mu_j$ (λ_j) at node i and node j, respectively. In the following discussion, $U(r, \epsilon)$ is the function defined as

$$U(r, \epsilon) = \arg\min_k \left\{ \left(\sum_{i=0}^{k} e^{-r} \frac{r^i}{i!} \right) > 1 - \epsilon \right\},$$

so that, if γ is a Poisson distributed random variable with parameter r then $\Pr\{\gamma > U(r, \epsilon)\} < \epsilon$.

First of all, it is observed that the existence of the reflecting boundary, $f_i = 0$ affects the value of $F_\tau(*)$ because the departure rate from node i is 0 along this line. However, as discussed in the following lemma, this effect gets smaller with larger f_i^* and can be neglected when f_i^* is greater than D_i given below.

$$D_i = U(\tau\mu_i(w_i^* - 1), \epsilon). \tag{5.9}$$

Lemma 4. *Let p_B be the probability of the event of hitting the reflecting boundary at $f_i = 0$ before the first passage to D occurs starting from c. Then for a given $\epsilon > 0$, $p_B < \epsilon$ if $f_i^* > D_i = U(\tau\mu_i(w_i^* - 1), \epsilon)$.*

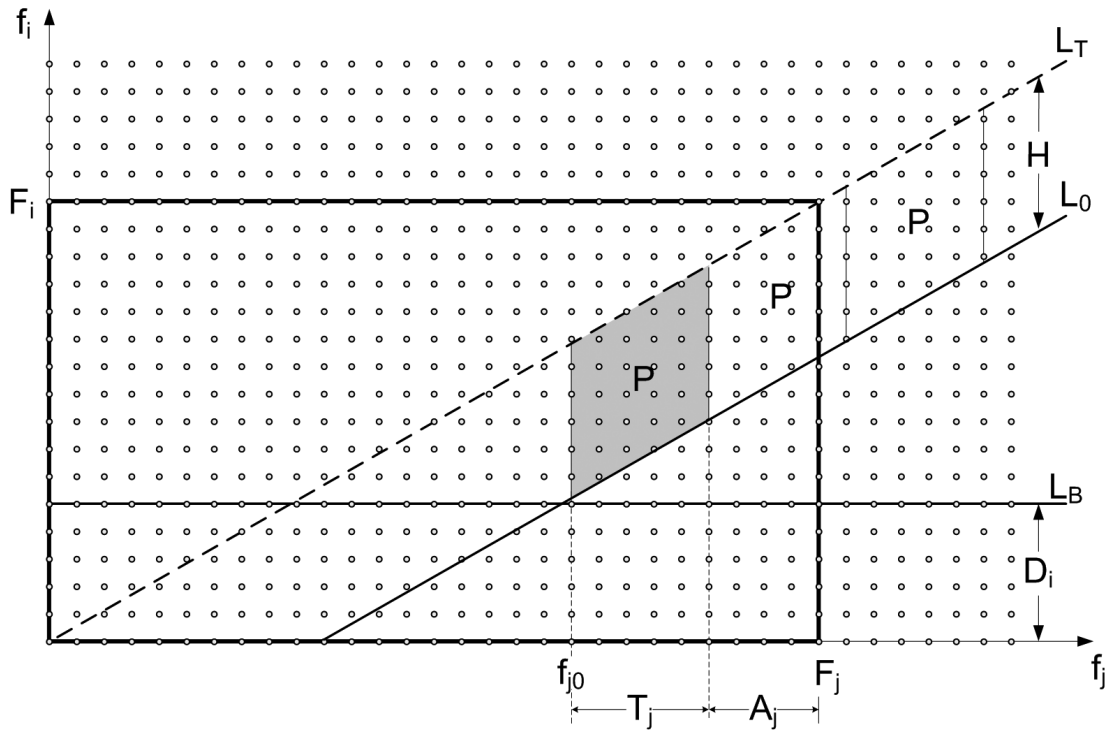


Figure 5.5: Calculation of $F_\tau(*)$.

Proof. Let B denote the event that the boundary $f_i = 0$ is hit in time interval $0 \leq u < \tau$ and B' its complementary event. The probability of event B can be written as

$$p_B = \Pr\{B\} = 1 - \Pr\{f_i(u) > 0, 0 \leq u < \tau\}.$$

p_B can be bounded as follows

$$\begin{aligned} p_B &= 1 - \Pr\{f_i^* + a_i(u) - d_i(u) > 0, 0 \leq u < \tau\} \\ &< 1 - \Pr\{f_i^* - d_i(u) > 0, 0 \leq u < \tau\} \\ &= 1 - \Pr\{d_i(\tau) < f_i^*\} \\ &< 1 - \Pr\left\{\hat{d}_i(\tau) < f_i^*\right\} \end{aligned}$$

where $\hat{d}_i$ is a Poisson random variable with parameter $\tau\mu_i(w_i - 1)$. The last line follows from the fact that $d_i(\tau)$ is not exactly Poisson since the departure rate at $f_i = 0$ is 0. Hence, p_B converges to 0 as f_i^* increases. If $f_i^* > D_i$, then

$$p_B < 1 - \Pr\left\{\hat{d}_i(\tau) < D_i\right\} < 1 - (1 - \epsilon) = \epsilon$$

□

The following lemma shows that, if $f_i^* > D_i$ then $F_\tau(*)$ becomes approximately a function of the distance between the point c and the line L_T, which can be calculated as

$$h((f_i^*, f_j^*)) = m^* f_j^* - f_i^*. \tag{5.10}$$

Lemma 5. *If $f_i^* > D_i$, then $F_\tau(*)$ is approximately a function of $h((f_i^*, f_j^*)) = m^* f_j^* - f_i^*$, where $h((f_i^*, f_j^*))$ corresponds to the vertical distance between c and the line L_T given by the equation $f_i = m^* f_j$.*

Proof. $F_\tau(*)$ can be decomposed into two terms conditioned on B:

$$\begin{aligned} F_\tau(*) &= \Pr\{T_{cD} < \tau\} \\ &= p_B \Pr\{T_{cD} < \tau \mid B\} + (1 - P_B) \Pr\{T_{cD} < \tau \mid B'\} \end{aligned}$$

If $f_i^* > D_i$, then using Lemma 4 $F_\tau(*)$ can be approximated as:

$$\begin{aligned} F_\tau(*) &\approx \Pr\{T_{cD} < \tau \mid B'\} \\ &= Pr\{\inf\{u > 0 : h(f_i(u), f_j(u)) < 0\}\} \end{aligned}$$

where

$$\begin{aligned} h(f_i(u), f_j(u)) &= m^*(f_j^* + a_j(u) - d_j(u)) - (f_i^* + a_i(u) - d_i(u)) \\ &= (m^* f_j^* - f_i^*) - m^*(d_j(u) - a_j(u)) + (d_i(u) - a_i(u)) \\ &= h(f_i^*, f_j^*) - m^*(d_j(u) - a_j(u)) + (d_i(u) - a_i(u)) \end{aligned}$$

Since, a_i, d_i, a_j, and d_j are independent Poisson processes with state independent rates, $F_\tau(*)$ is a function of $h(f_i^*, f_j^*)$. □

It is also possible to set $F_\tau(*) = 0$ if c is sufficiently far from L_T. The following lemma states that this condition is satisfied when $h((f_i^*, f_j^*)) > H$, where H is given by

$$H = m^* D_j + A_i, \tag{5.11}$$

where $A_i = U(\tau\lambda_i, \epsilon)$, and $D_j = U(\tau\mu_j w_j, \epsilon)$.

Lemma 6. *If $h((f_i^*, f_j^*)) > H$, then $F_\tau(*)$ is smaller than 2ϵ, where $H = m^* D_j + A_i$, $A_i = U(\tau\lambda_i, \epsilon)$, and $D_j = U(\tau\mu_j w_j, \epsilon)$.*

Proof. $F_\tau(*)$ can be partitioned conditioning on the number of arrivals and departures at nodes i and j during the time interval τ. Using $P_x(y)$ as a shorthand notation for $\Pr\{x = y\}$,

$$\begin{aligned}
F_\tau(*) &= \sum_{i^+=0}^{\infty}\sum_{i^-=0}^{\infty}\sum_{j^+=0}^{\infty}\sum_{j^-=0}^{\infty} P_{a_i}(i^+)P_{d_i}(i^-)P_{a_j}(j^+)P_{d_j}(j^-) \\
&\quad \Pr\left\{T_{cD} < \tau \mid a_i = i^+, d_i = i^-, a_j = j^+, d_j = j^-\right\} \\
&< \sum_{i^+=0}^{\infty}\sum_{j^-=0}^{\infty} P_{a_i}(i^+)P_{n_j}(j^-)\Pr\left\{T_{cD} < \tau \mid a_i = i^+, d_j = j^-\right\} \\
&= \left(\sum_{i^+=0}^{A_i}\sum_{j^-=0}^{D_j} + \sum_{i^+=0}^{\infty}\sum_{j^-=D_j}^{\infty} + \sum_{i^+=A_i}^{\infty}\sum_{j^-=0}^{\infty} - \sum_{i^+=A_i}^{\infty}\sum_{j^-=D_j}^{\infty}\right) \\
&\quad P_{a_i}(i^+)P_{n_j}(j^-)\Pr\left\{T_{cD} < \tau \mid a_i = i^+, d_j = j^-\right\} \\
&< \sum_{i^+=0}^{A_i}\sum_{j^-=0}^{D_j} P_{a_i}(i^+)P_{n_j}(j^-)\Pr\left\{T_{cD} < \tau \mid a_i = i^+, d_j = j^-\right\} + \\
&\quad \epsilon + \epsilon - \epsilon^2
\end{aligned}$$

Observe that if $m^*(f_j^* - D_j) - (f_i^* + A_i) > 0$ then the first term is 0, and $F_\tau(*) < (2\epsilon - \epsilon^2) < 2\epsilon$. □

With this assumption, if c is below the line L_0 then $F_\tau(*)$ can be approximated as 0. Therefore, $F_\tau(*)$ is non-zero only if c lies between the lines L_T and L_0 shown in Figure 5.5.

If the slope of L_T is approximated as

$$m^* \approx (w_i^* - 0.5)/(w_j^* + 0.5), \tag{5.12}$$

then it suffices to calculate $F_\tau(*)$ for c in the shaded region, P. Because, P is repeating itself along the strip between the lines L_T and L_0. The period in f_j dimension is given by

$$T_j = (2w_i^* - 1)/\gcd(2w_i^* - 1, 2w_j^* + 1) \tag{5.13}$$

and the corresponding period in f_j dimension is

$$T_j = (2w_i^* - 1)/\gcd(2w_i^* - 1, 2w_j^* + 1), \tag{5.14}$$

as stated by the following lemma. Hence all of the states in this strip can be mapped to a state in P.

Lemma 7. *Let* $T_j = (2w_i^* - 1)/\gcd(2w_i^* - 1, 2w_j^* + 1)$ *and* $T_i = m^* T_j$, *where* $gcd(x, y)$ *denotes the greatest common divisor of numbers* x *and* y. $F_\tau(*)$ *for* $c = (f_i^* nT_i, f_j^* + nT_j)$ *for any integer* $n > 0$ *is equal to the* $F_\tau(*)$ *for* $c = (f_i^*, f_j^*)$ *if* $f_i^* > D_i$.

Proof. T_j is an integer by the definition of gcd. Since w_i^* and w_j^* are integers $\gcd(2w_i^* - 1, 2w_j^* + 1)$ is an integer. Therefore,

$$\begin{aligned} T_i &= m^* T_j \\ &= \frac{w_i^* - 0.5}{w_j^* + 0.5} \frac{2w_j^* + 1}{\gcd(2w_i^* - 1, 2w_j^* + 1)} \\ &= \frac{2w_i^* - 1}{\gcd(2w_i^* - 1, 2w_j^* + 1)} \end{aligned}$$

is also an integer. Then,

$$\begin{aligned} h(f_i^* + nT_i, f_j^* + nT_j) &= m^*(f_j^* + nT_j) - (f_i^* + nT_i) \\ &= m^* f_j^* - f_i^* + m^* nT_j - nT_i \\ &= h(f_i^*, f_j^*) \end{aligned}$$

If $f_i^* > F_i$ then due to Lemma 5 $F_\tau(*)$ can be approximated as a function of $h(f_i^*, f_j^*)$. Since, $h(f_i^* + nT_i, f_j^* + nT_j) = h(f_i^*, f_j^*)$, $F_\tau(*)$ for these states are equal. □

Therefore, for a given wavelength allocation w_i^* and w_j^*, $F_\tau(*)$ for any $c = (f_i^*, f_j^*)$ can be obtained by using the values $F_\tau *$ calculated considering only the states $0 \leq f_j \leq F_j$ and $0 \leq f_i \leq F_i$. F_j can be calculated as

$$F_j = f_{j0} + T_j + A_j, \tag{5.15}$$

where

$$f_{j0} = \lceil (D_i + H)/m^* \rceil, \tag{5.16}$$

$$A_j = U(\tau\lambda_j, \epsilon) \tag{5.17}$$

where A_j is a margin added so that the effects of the boundary on the right hand side can be ignored. T_j, D_i, H, and m^* are given by equations (5.13), (5.9),(5.11), and (5.3) respectively. Finally, F_i can be taken as

$$F_i = \lceil m^* F_j \rceil \tag{5.18}$$

To sum up, for the efficient calculation of $F_\tau(*)$, first the values are calculated for a representative set of states and recorded. Then, this data is used to estimate $F_\tau(*)$ for any state required.

To obtain the $F_\tau(*)$ for a representative set of states, F_i and F_j are calculated using (5.18) and (5.15), respectively. Then the number of flows at node i and node j are truncated at F_i and F_j, respectively, as discussed in Note 5.3.1. On the resulting finite two-dimensional chain the the set of equations given in 5.8 is solved to yield $F_\tau(*)$ values for all states c with $f_i^* = 0, \ldots, F_i$ and $f_j^* = 0, \ldots, F_j$. The outline of this calculation is presented in Algorithm 1, where sample values of $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ are calculated along with the parameters T_i, T_j and f_{j0} for each possible combination of $\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*$.

Using Lemma 6 and Lemma 7, the value of $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ for any $c = (f_i^*, f_j^*)$ can be obtained based on the sample values of $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ that are calculated by Algorithm 1 for the corresponding values of $\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*$. The steps for this calculation are shown in Algorithm 2.

5.3.4 Computational Complexity and Storage Requirements for the HM3 Method

HM3 algorithm requires the computation of $\{v_{ij}\}$ values for each valid action a_{ij} at each decision instant. Noting that there may be at most $N \times (N-1)$ valid actions at any time, $\{v_{ij}\}$ values are calculated using the Algorithm 2 which is basically a table look-up operation. Therefore, the on-line part of the HM3 algorithm requires at most $N \times (N-1)$ table look-ups.

Algorithm 1 Calculation of the sample values for $F_\tau()$.

Input: N, W, τ, ϵ, and λ_i, μ_i
Output: $M_F, M_{T_i}, M_{T_j}, M_{f_{j0}}$

for all possible combinations of $(\lambda_i, \lambda_j, \mu_i, \mu_j)$ **do**
  for $w_i^* = 2$ to W_{max} **do**
    for $w_j^* = 1$ to $(W_{max} + 1 - w_i^*)$ **do**
      1: $M_{T_j}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*] \leftarrow T_j$, given by (5.13).
      2: $M_{T_i}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*] \leftarrow T_i$, calculated using (5.14).
      3: $M_{f_{j0}}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*] \leftarrow f_{j0}$, obtained using (5.16).
      4: Calculate F_j and F_i using (5.15) and (5.18), respectively.
      5: Obtain $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ for $f_i^* = 1, \ldots, F_i$, $f_j^* = 1, \ldots, F_j$ as described in Section 5.3.2.
      6: $M_F[f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*] \leftarrow F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$, for $f_i^* = 1, \ldots, F_i$, $f_j^* = 1, \ldots, F_j$.
    end for
  end for
end for

Algorithm 2 Calculation of $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$ from sample values.

Input: $f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*, M_F, M_{T_i}, M_{T_j}, M_{f_{j0}}$
Output: $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*)$

if $(f_i^*/f_j^*) > m^*$, given in (5.12) **then**
  $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*) \leftarrow 1$
else if $h(f_i^*, f_j^*) > H$, where $h(f_i^*, f_j^*)$ and H are defined by (5.10) and (5.11), **then**
  $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*) \leftarrow 0$
else
  $T_i \leftarrow M_{T_i}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*]$
  $T_j \leftarrow M_{T_j}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*]$
  $f_{j0} \leftarrow M_{f_{j0}}[\lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*]$
  $c \leftarrow \max(0, \lfloor (f_j^* - f_{j0})/T_j \rfloor)$
  $\tilde{f}_i \leftarrow f_i^* - cT_i$
  $\tilde{f}_j \leftarrow f_j^* - cT_j$
  $F_\tau(f_i^*, f_j^*, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*) \leftarrow M_F[\tilde{f}_i, \tilde{f}_j, \lambda_i, \lambda_j, \mu_i, \mu_j, w_i^*, w_j^*]$
end if

Computationally intense part of HM3 method is the off-line calculations presented in Algorithm 1. For each possible combination of the $(\lambda_i, \lambda_j, \mu_i, \mu_j) \times (w_i, w_j)$, the set of linear equations given in (5.8) is solved. Let C denote the number of possible combinations of $(\lambda_i, \lambda_j, \mu_i, \mu_j)$. The number of possible combinations of (w_i, w_j) is smaller than $(W - N + 1)^2$, since at most $(W - N + 1)$ wavelengths can be allocated to a node due to the connectivity constraint. Then, there are at most $C \times (W-N+1)^2$ possible combinations of $(\lambda_i, \lambda_j, \mu_i, \mu_j) \times (w_i, w_j)$ and (5.8) is solved at most $C \times (W-N+1)^2$ times. The computational complexity of solving (5.8) is n^3, where $n = F_i \times F_j$ is the number of equations. Denoting the maximum values of F_i and F_j as F_i^{max} and F_j^{max}, respectively, the computational complexity of HM3 is found to be $C \times (W - N + 1)^2 \times (F_i^{max} \times F_j^{max})^3$. The storage requirement for HM3 is equal to $C \times (W - N + 1)^2 \times (F_i^{max} \times F_j^{max})$. The values of F_i^{max} and F_j^{max} are calculated below.

Since, $w_i^* \leq (W - N + 1)$ and $gcd(x, y) \geq 1$, T_j given in (5.13) satisfies

$$T_j \leq 2(W - N + 1) - 1. \tag{5.19}$$

m^* defined in (5.12) can be lower bounded as

$$m^* = \frac{w_i^* - 0.5}{w_j^* + 0.5} \geq \frac{0.5}{W - N + 1 + 0.5} > \frac{1}{2W} \tag{5.20}$$

An upper bound for the f_{j0} parameter given in (5.16) is obtained as follows.

$$\begin{aligned} f_{j0} &= \left\lceil \frac{U(\tau\mu_i(w_i^* - 1), \epsilon) + m^* U(\tau\mu_j w_j, \epsilon) + U(\tau\lambda_i, \epsilon)}{m^*} \right\rceil \\ &= \left\lceil U(\tau\mu_j w_j, \epsilon) + \frac{U(\tau\mu_i(w_i^* - 1), \epsilon) + U(\tau\lambda_i, \epsilon)}{m^*} \right\rceil \end{aligned}$$

Using (5.20),

$$f_{j0} \leq \left\lceil U\left(\tau\mu_{max}(W - N + 1), \epsilon\right) + \frac{U(\tau\mu_{max}(W - N), \epsilon) + U(\tau\lambda_{max}, \epsilon)}{m^*} \right\rceil \tag{5.21}$$

$$f_{j0} \leq U\left(\tau\mu_{max}(W - N + 1), \epsilon\right) + \frac{U(\tau\mu_{max}(W - N), \epsilon) + U(\tau\lambda_{max}, \epsilon)}{1/2W} + 1 \tag{5.22}$$

where λ_{max} and μ_{max} are upper bounds for the arrival and departure rates, respectively. With this definitions, A_j given in (5.17) satisfies

$$A_j \leq U(\tau\lambda_{max}, \epsilon) \tag{5.23}$$

Using (5.19), (5.22) and (5.23), F_j defined in 5.15 can be bounded as

$$\begin{aligned} F_j < F_j^{max} &= U\left(\tau\mu_{max}(W-N+1),\epsilon\right) + \frac{U(\tau\mu_{max}(W-N),\epsilon) + U(\tau\lambda_{max},\epsilon)}{1/2W} + \\ & \quad 2(W-N+1) + U(\tau\lambda_{max},\epsilon) \end{aligned}$$

Upper bound for m^* can be calculated as

$$m^* = \frac{w_i^* - 0.5}{w_j^* + 0.5} \leq \frac{W-N+1-0.5}{1+0.5} \leq \frac{W-N+0.5}{1.5} \tag{5.24}$$

Using (5.18), (5.21), (5.19), (5.23) and (5.24),

$$\begin{aligned} F_i < F_i^{max} &= U(\tau\mu_{max}(W-N),\epsilon) + U(\tau\lambda_{max},\epsilon) + \\ & \quad \frac{W-N+0.5}{1.5}\left\{U\left(\tau\mu_{max}(W-N+1),\epsilon\right) + 1 + \right. \\ & \quad \left. U(\tau\lambda_{max},\epsilon) + 2(W-N+1)\right\} \end{aligned}$$

As an example, following values are obtained for the 3-node test network used in Section 4.5, with $\lambda = 0.5$ and $\epsilon = 10^{-5}$.

$$\begin{aligned} N &= 3 & U\left(\tau\mu_{max}(W-N+1),\epsilon\right) &= 4 \\ W &= 7 & U\left(\tau\mu_{max}(W-N),\epsilon\right) &= 4 \\ \lambda_{max} &= 2 & U(\tau\lambda_{max},\epsilon) &= 3 \\ \mu_{max} &= 1 & C &= 3 \\ \tau &= 0.05 & F_j^{max} &= 115 \\ & & F_i^{max} &= 61 \end{aligned}$$

As a second example, a larger network with 10 nodes and 30 wavelengths is considered. Assuming that the arrival and departure rates at each node are distinct, $\lambda_{max} = 2$ and $\mu_{max} = 1$, the results obtained are:

$$\begin{aligned} N &= 10 & U\left(\tau\mu_{max}(W-N+1),\epsilon\right) &= 8 \\ W &= 30 & U\left(\tau\mu_{max}(W-N),\epsilon\right) &= 8 \\ \lambda_{max} &= 2 & U(\tau\lambda_{max},\epsilon) &= 3 \\ \mu_{max} &= 1 & C &= 45 \\ \tau &= 0.05 & F_j^{max} &= 713 \\ & & F_i^{max} &= 767 \end{aligned}$$

Chapter 6

Performance of the Heuristic Methods

In order to evaluate the performance of heuristic DWA methods developed in Chapter 5, simulations are performed under a wide range of experimental settings. In this chapter, heuristic methods are compared with each other and with the exact solutions obtained in Chapter 4, based on the performance measures discussed in Section 3.4.1.

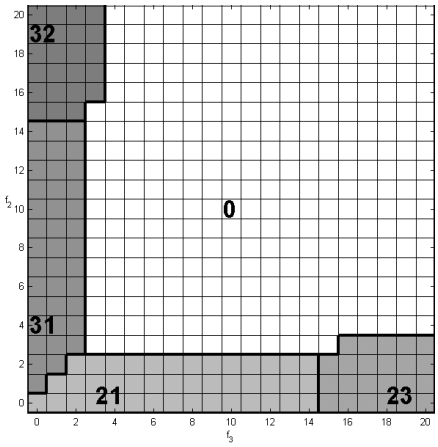
First, the case of stationary traffic is considered and the efficiency of heuristic methods are demonstrated for a range of flow arrival rates. These results are used to compare the relative performance of each method at different network load levels. Temporal characteristics of DWA methods are inspected using the resulting wavelength distributions in time. Afterwards, the effects of V_{thr} parameter on the performance of HM3 method is analyzed. Heuristic methods are also tested under non-stationary traffic patterns, using time varying arrival rates at each node. Moreover, the effects of average flow size, channel bandwidth, average reconfiguration delay and number of wavelengths on the performance of the DWA methods are investigated.

6.1 Stationary Traffic

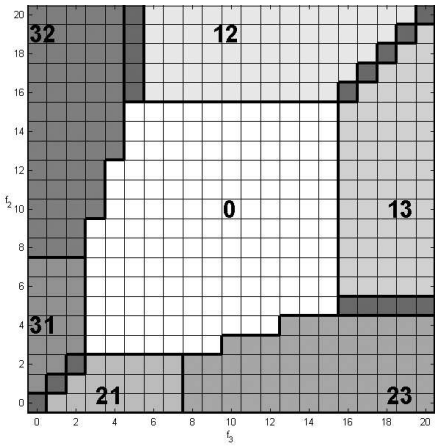
In the stationary traffic case, it is assumed that the average flow arrival rates at each node do not change in time. To be able to compare results with the exact solutions obtained in Chapter 4, the 3-node test network given in Figure 4.4 is used with the same traffic pattern. Namely, number of wavelengths $W = 7$, service rate of a flow by a single wavelength channel $\mu_i = 1$ flows/s at each node i and average reconfiguration delay, $(1/\sigma) = 50$ ms. Relative arrival rates are λ, 2λ and 4λ at node 1, node 2 and node 3, respectively.

Heuristic methods discussed in Chapter 5 are used to obtain reconfiguration policies for $\lambda = 0.7$. The exact policies corresponding to this case are obtained in Chapter 4 and parts of the policies corresponding to a sample set of states are given in Figure 4.5. The parts of the heuristic policies corresponding to the same set of states are provided in Figure 6.1 for comparison purposes. It is observed that HM1 and FS policies have similar behavior. Both of them make switching actions when the state is close to the axes, i.e., when the number of flows is small at a node. The absence of actions a_{12} and a_{13} in the matrices shows that these policies do not reconfigure the wavelengths to balance the load in the network. Unlike the corresponding exact NFS policy, HM2 results in a symmetric matrix because it does not consider the arrival rates, which are inhomogeneous for the test network. Due to this symmetric nature, there are states at which more than one action have the same cost. These states are shown unlabeled in the figure. It is also observed that HM2 makes switches from node 1 when the number of flows at other nodes gets large. Therefore, it may be concluded that HM2 makes switches to balance the load. Compared to the NFS policy, HM2 performs reconfiguration at a larger number of states. Finally, Figure 6.1c plots the policy obtained with HM3 using $V_{thr} = 0.85$. Similar to the NSFS policy, switches from node 1 are performed when the number of flows at node 2 or node 3 gets large. The area corresponding to no-switching decisions is narrower for HM3 compared to HM1 and HM2. Hence, HM3 performs switches at a larger number of states and the resulting policy closely resembles the NSFS policy.

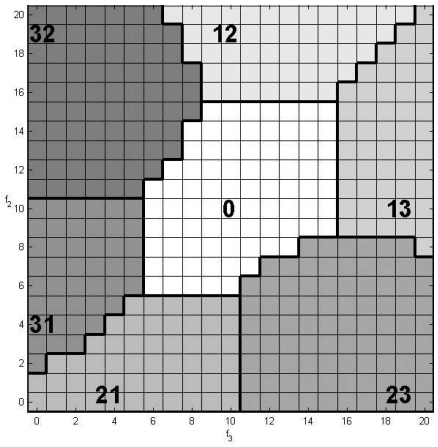
To evaluate the performance of heuristic policies at different network load



(a) Heuristic policy HM1



(b) Heuristic policy HM2



(c) Heuristic policy HM3

Figure 6.1: Heuristic switching policies for the 3-node test network, for states with $w = [3, 2, 2]$ and $f = [15, f_2, f_3]$.

levels, simulations are performed for several λ values and results are plotted in Figure 6.2 as a function of λ. Corresponding graphs obtained with the optimum policies are shown in Figure 4.6. It is observed that, on the average HM1 performs 10% worse then FS in terms of slowdown. HM1 achieves holding cost values which are nearly 5% higher than FS and the fairness of HM1 and FS policies are comparable. With respect to FS, HM1 is more conservative in making reconfigurations and results in less number of switches. HM2 obtains slowdown values which are close to exact policy NFS. The difference is always below 10% and becomes much smaller at moderate load levels. The same conclusion applies also for the holding cost. The fairness of HM2 is better than NFS up to $\lambda = 0.5$

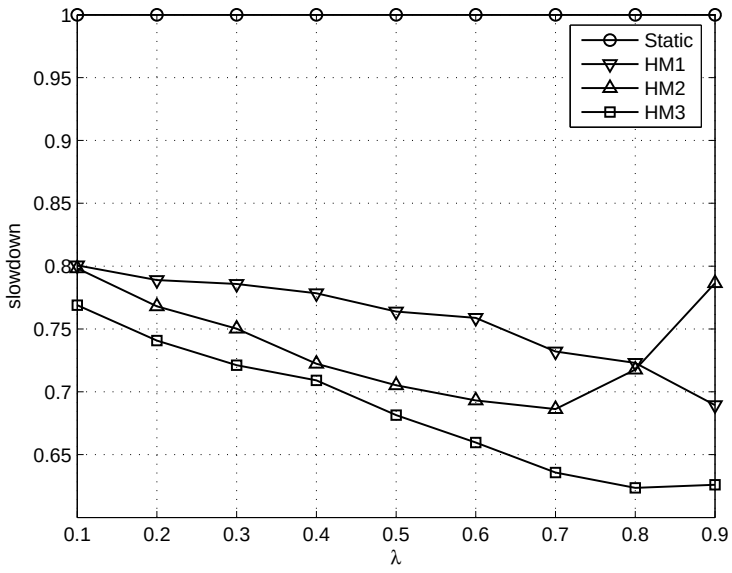
and HM2 makes less number of reconfigurations. Finally, when HM3 is compared with NSFS the following are observed. In terms of slowdown and holding cost, HM3 obtains nearly 5% worse results than the optimum policy. However, as λ increases the optimality gap closes and the difference is as low as 1.5% for high levels of network load. Moreover, HM3 achieves higher fairness and makes much less wavelength switching compared to NSFS.

When heuristic methods are compared to each other, it can be concluded that HM3 achieves best slowdown and fairness results by making a moderate number of reconfigurations at any network load level. HM2 policy has a good slowdown and holding cost performance only for small values of network load. The performance of HM2 gets significantly worse as the network load is increased. It makes the largest number of reconfigurations and attains a good level of fairness for all values of λ. On the other hand, HM1 performs well in terms of holding cost. However, it has the worst slowdown and fairness curves among the heuristic methods.

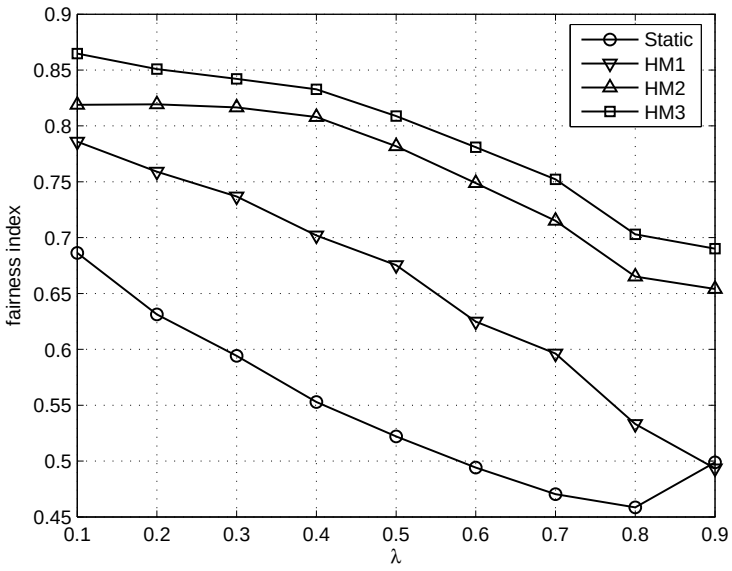
The heuristic methods can also be compared based on their ability to achieve load balance by inspecting their time domain behavior. For this purpose, the degree of load imbalance at any time can be measured as the distance between the wavelength allocation vector $\boldsymbol{w}$ and the ideal allocation vector $\hat{\boldsymbol{w}} = [\hat{w}_i]$, where $\hat{w}_i$ is defined as $\hat{w}_i = (f_i / \sum_i f_i)\, W$. Hence, $\hat{w}_i$ is the amount of bandwidth (in terms of wavelength channels) that node i should be allocated in order to achieve a perfect load balance. It is ideal because granularity constraint is not considered and it may take values smaller than 1, in contrast to the case in the DWA problem where w_i is an integer and $w_i \geq 1$. The load imbalance factor, I_w, is then defined as

$$I_w = |\boldsymbol{w} - \hat{\boldsymbol{w}}| = \sqrt{\sum_i \left(w_i - \frac{f_i}{\sum_j f_j} W \right)^2} \tag{6.1}$$

To compare the load balancing behavior of the heuristic methods, simulations are performed with each method using identical traffic patterns and setting $\lambda = 0.8$. Load imbalance is calculated using (6.1) and smoothed using a moving average filter with a window width of 20 seconds. The results are plotted in Figure 6.3. Load imbalance of static wavelength allocation takes values as high as

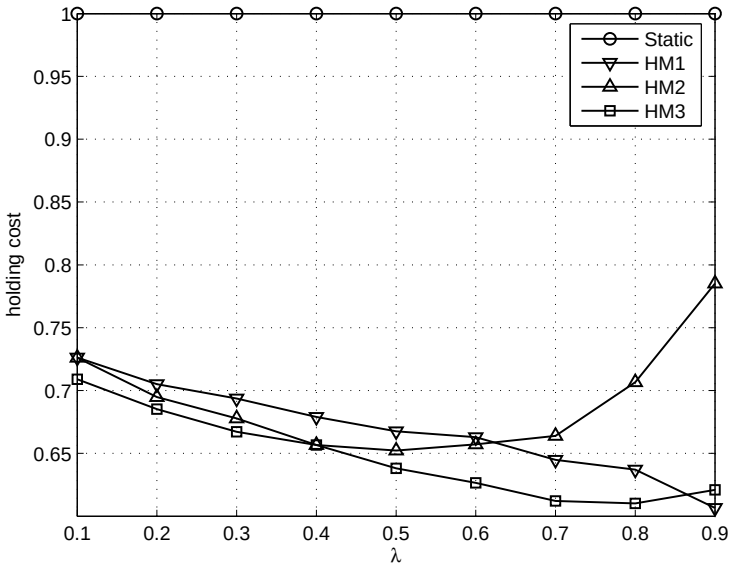


(a) Slowdown

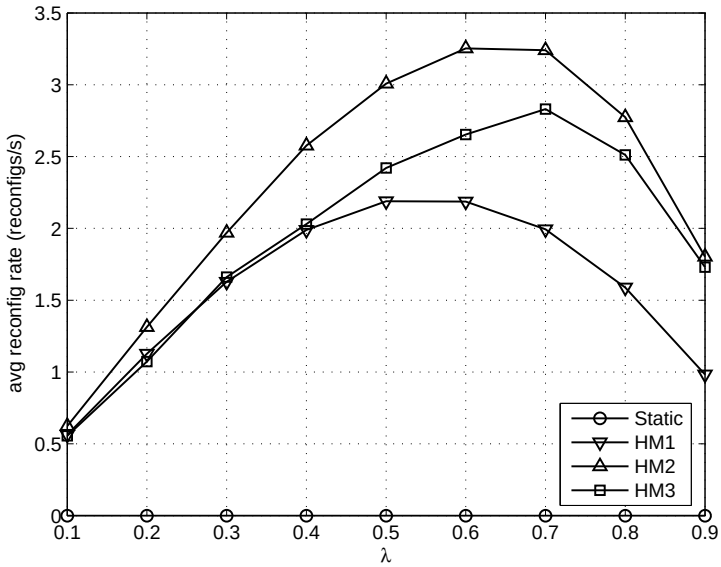


(b) Fairness

Figure 6.2: Performance of heuristic methods as a function of network load.



(c) Holding Cost



(d) Average Switching Rate

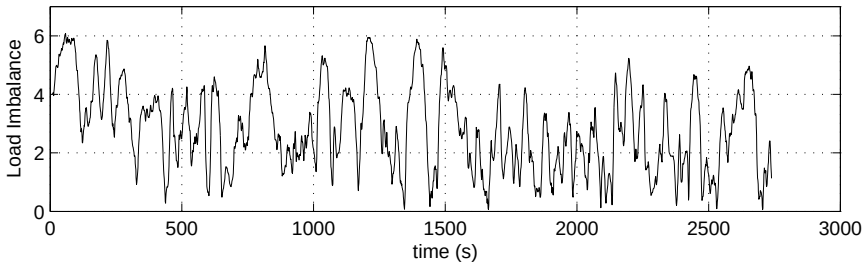
Figure 6.2 (continued): Performance of heuristic methods as a function of network load.

6 and is the worst with an average value of 2.8143. The second worst performance belongs to HM2 with I_w usually in the interval [1,2]. Time averaged value of load imbalance is calculated to be 1.4081 for HM2. HM1 has a better load balancing behavior. I_w varies in [0,2] and shows larger fluctuations. The average value of load imbalance with HM1 is 0.7412. Finally, HM3 results in load imbalance which is smaller than 1 most of the time. It has the minimum average I_w, which is found to be 0.6138. This value corresponds to nearly 20% of the load imbalance of the static allocation. The enhancement in I_w is 60% and 20% with respect to HM2 and HM1, respectively.

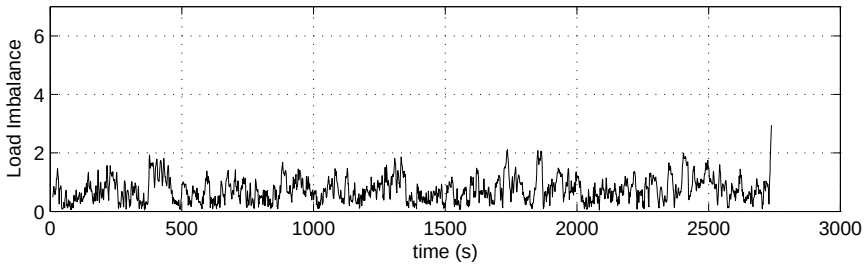
Finally, the number of wavelengths allocated to each node in time are plotted in Figure 6.4 for $\lambda = 0.1$. The data is smoothed using a moving average filter of length 20 s. to increase readability. The optimum SWA is calculated according to Section 3.4.5 and plotted as a horizontal line in the figure corresponding to the related node. The optimum SWA is $\bar{\boldsymbol{w}} = [1.53, 2.22, 3.25]$ and it is observed that the DWA results in a wavelength distribution which takes values around the optimum SWA. With DWA, the overall time average of wavelengths at each node is $\boldsymbol{w_{avg}} = [1.43, 2.24, 3.30]$. The small difference between $\boldsymbol{w_{avg}}$ and $\bar{\boldsymbol{w}}$ is due to the connectivity constraint which prevents assignment of 0 wavelengths to a node. Moreover, $\boldsymbol{w_{avg}}$ sum to 6.97 due to the unavailability of wavelengths during reconfiguration periods.

Similarly, Figure 6.5 plots the results for $\lambda = 0.9$. Optimum SWA for this case is $\bar{\boldsymbol{w}} = [1.06, 2.02, 3.92]$ and overall average of wavelengths assigned to each node under DWA is $\boldsymbol{w_{avg}} = [1.23, 1.94, 3.73]$. Comparing Figure 6.4 and Figure 6.5 it can be observed that as the arrival rate is increased the wavelength allocation experiences less variability. For $\lambda = 0.9$, $\boldsymbol{w_{avg}}$ sum up to 6.91, which indicates that more reconfigurations are performed with respect to $\lambda = 0.1$ case.

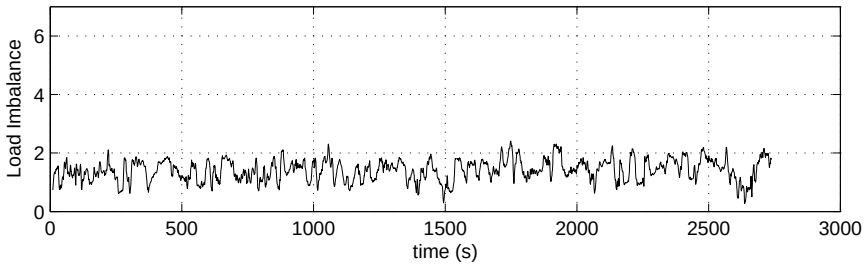
The results presented in this section, clearly demonstrate that dynamic wavelength allocation can improve the network performance even in the case of stationary traffic. Moreover, it has also been observed that a poor DWA policy can have dramatic results and perform even worse than static wavelength allocation. In the next section, the heuristic methods are evaluated under non-stationary



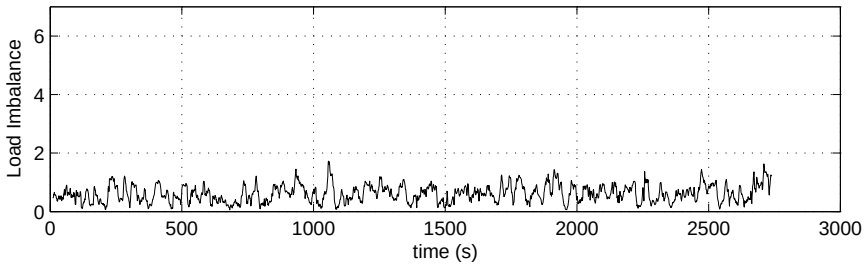
(a) Static



(b) HM1



(c) HM2



(d) HM3

Figure 6.3: Temporal behavior of load imbalance experienced by the wavelength allocation policies under stationary traffic.

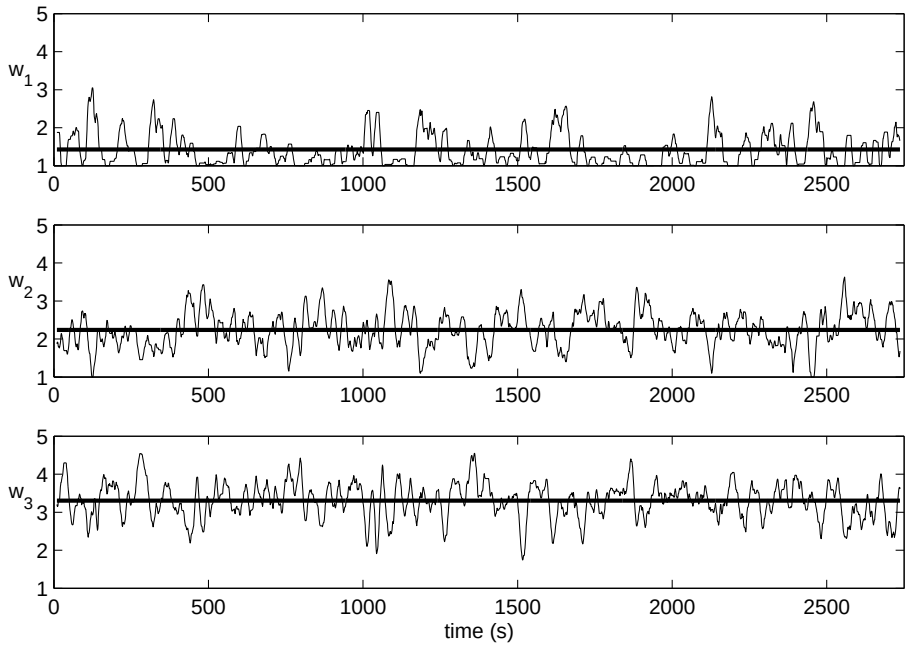


Figure 6.4: Average number of wavelengths at each node for $\lambda = 0.1$.

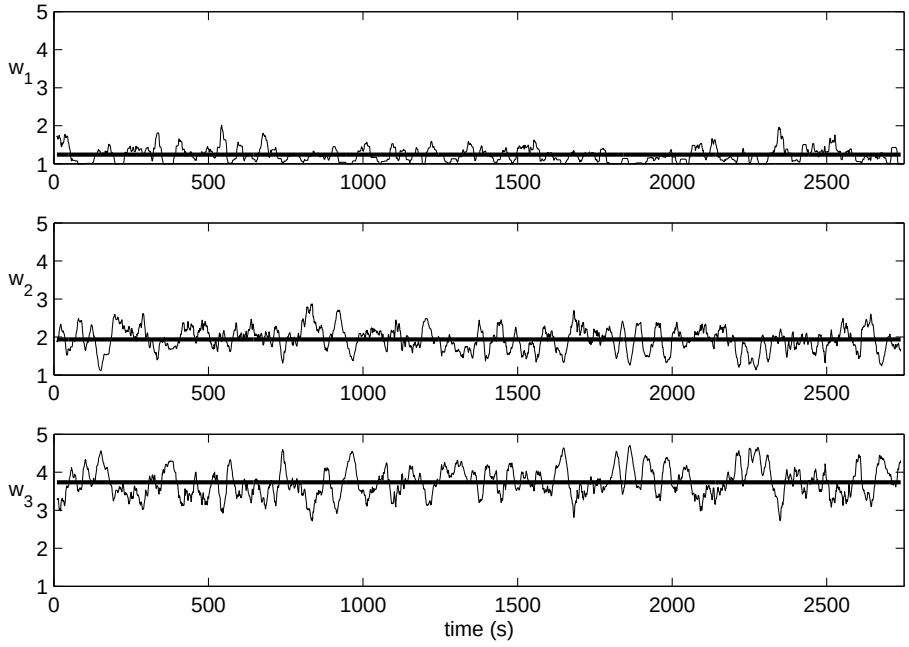


Figure 6.5: Average number of wavelengths at each node for $\lambda = 0.9$.

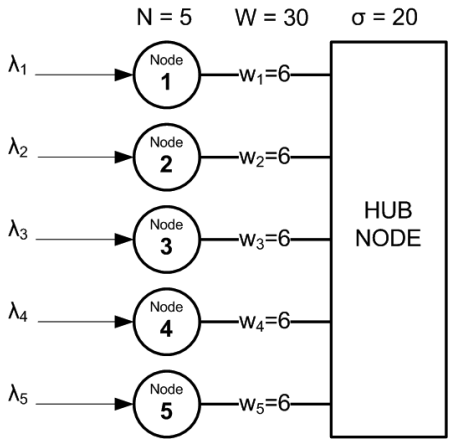


Figure 6.6: 5-node test network.

traffic demand where flow arrival rates exhibit variations as a function of time.

6.2 Non-stationary Traffic

As discussed in Section 3.2, metro access networks have traffic patterns changing in time. Hence, proper adaptability of the DWA methods to time varying traffic characteristics is an important requirement. To test the performance of the heuristic methods under non-stationary traffic conditions, simulations are performed using the network shown in Figure 6.6, which has $N = 5$ nodes and $W = 30$ wavelengths. μ_i is 1 flows/s at each node i and average reconfiguration delay is 50 ms. Flow arrival rates are changed in time according to Table 6.1.

Since the time average of flow arrival rates at the nodes are equal, 6 wavelength channels are allocated to each node with the static policy. Performance metrics are calculated for the flows arriving in the time interval [500-2500] sec. and tabulated in Table 6.2. Overall network load is 0.5 at any time, and maximum load at any node is 0.83 with static wavelength allocation.

It is observed that HM1 succeeds to decrease the holding cost to less than 60% of the value obtained with static policy. However, the improvement in slowdown is below 30% and the enhancement in fairness is negligible. HM2 provides a much

Table 6.1: Time varying arrival rates.

time (s)	$\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_2}$	$\boldsymbol{\lambda_3}$	$\boldsymbol{\lambda_4}$	$\boldsymbol{\lambda_5}$
0– 500	3	3	3	3	3
500– 900	1	2	3	4	5
900–1300	2	3	4	5	1
1300–1700	3	4	5	1	2
1700–2100	4	5	1	2	3
2100–2500	5	1	2	3	4
2500–2750	3	3	3	3	3

Table 6.2: Comparison of heuristic policies under dynamic traffic conditions.

Method	**# Switch**	**Slowdown**	**Fairness**	**H. Cost**
Static	0	0.5737	0.4683	17186.0
HM1	21077	0.4119	0.4741	10017.0
HM2	23261	0.2958	0.6865	7840.9
HM3	17248	0.2679	0.7594	7345.3

better performance. It makes the highest number of reconfigurations among the dynamic policies and achieves nearly 50% lower slowdown and and 55% lower holding cost when compared to the static policy. Besides, it attains a considerable improvement in fairness. Finally, HM3 obtains the best performance metrics with a significantly less number of reconfigurations. When compared to HM2, HM3 is 9% better in terms of both holding cost and slowdown. With respect to the static policy, these numbers correspond to a decrease of 53% and 57% in slowdown and holding cost, respectively. The fact that HM3 achieves highest performance with lowest number of reconfigurations may indicate that HM3 performs the most beneficial actions and successfully eliminates unnecessary reconfigurations.

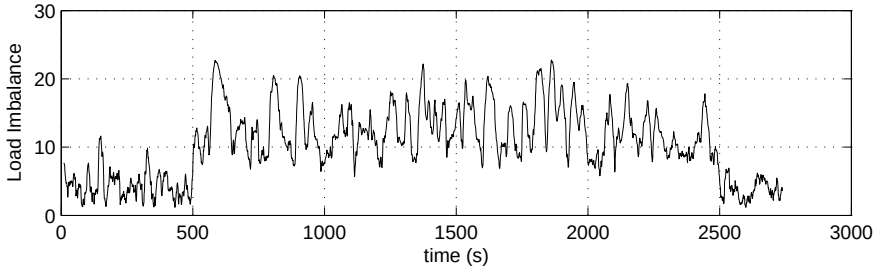
The load balance achieved with each policy is investigated using the load imbalance parameter, as in the Section 6.2. The results for the non-stationary traffic case are shown in Figure 6.7. As expected, the inefficiency of the static wavelength allocation scheme increases when the traffic at each node is unevenly distributed. The average value of L_w turns out to be 10.49 with static policy.

HM1 exhibits an interesting behavior. The results are better with respect to static policy in both periods of symmetric and asymmetric arrival rates. However, when the traffic arrival rates are not symmetric, the load imbalance increases about 5 times with respect to the load imbalance in the case of symmetric traffic demand. The average load imbalance is 7.5, which corresponds just a 25% improvement when compared to the static wavelength allocation. HM2 and HM3 clearly achieve relatively successful load balance performance. As indicated by their graphics, HM2 and HM3 successfully adapt to changing arrival rate conditions and can keep the load imbalance at fairly equal levels. Load imbalance does not differ between the periods of symmetric and asymmetric arrival rates. HM2 achieves an L_w value of 1.9594, which is less than 20% of the average load balance of static policy. With HM3, average of L_w further decrease to 1.4977, corresponding to a 25% improvement over HM2 or equivalently an average load imbalance which is just 14% of the static policy. These results also suggest that slowdown and fairness are indeed appropriate metrics to capture the load balance performance of DWA policies.

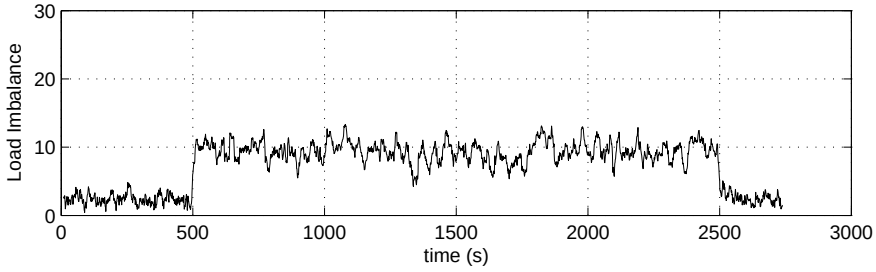
6.3 Sensitivity of HM3 Performance to V_{thr} Parameter

The parameter V_{thr} is the threshold used in HM3 algorithm in order to eliminate unnecessary switching actions. Conceptually, V_{thr} represents the value over which the long term expected rewards exceed the short term costs. Therefore, an action a_{ij} is applied only if v_{ij} corresponding to this action is larger than V_{thr}. If V_{thr} is set to 1, then no switching actions are performed since $v_{ij} < 1 \ \forall i, j$, and HM3 turns into a static wavelength allocation method. As the threshold is decreased, the frequency of wavelength switching increases and maximum number of reconfigurations are performed when $V_{thr} = 0$.

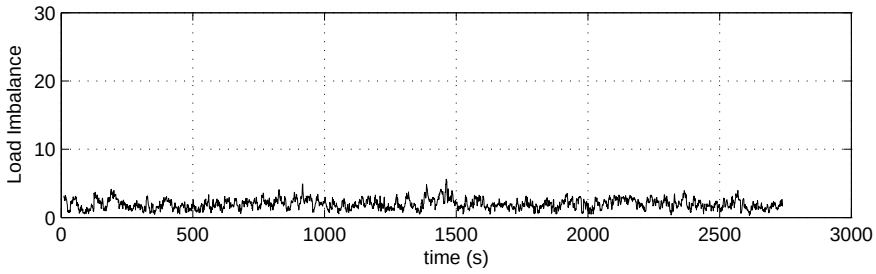
To demonstrate the effect of V_{thr} on the performance of the HM3 algorithm, experimental results are obtained on the 3-node test network used in Section 6.1. The value of the V_{thr} is set to 0.7, 0.85, 0.95 and 1, and the results are presented in



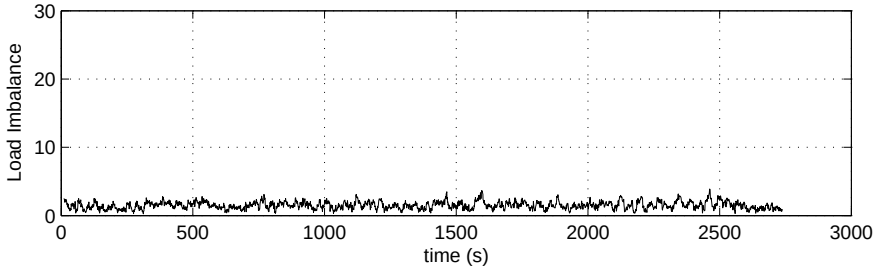
(a) Static



(b) HM1



(c) HM2



(d) HM3

Figure 6.7: Temporal behavior of load imbalance experienced by the wavelength allocation policies under non-stationary traffic.

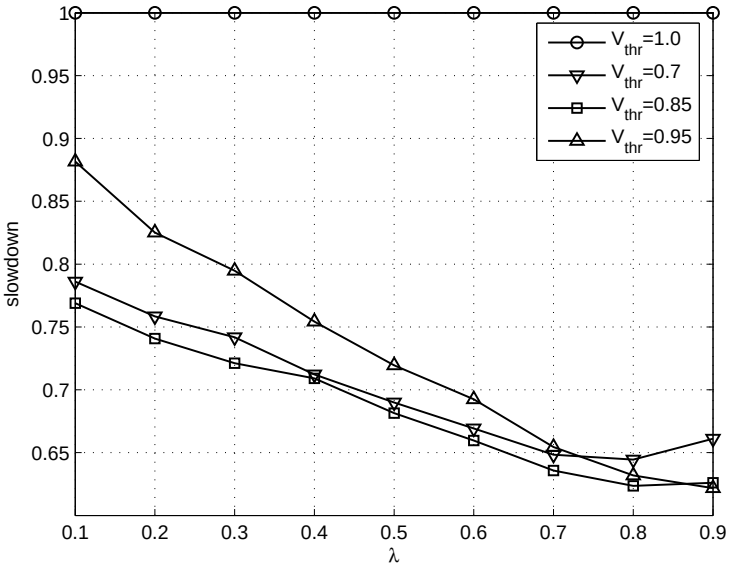
Figure 6.8. Slowdown obtained with each V_{thr} setting is normalized with respect to the slowdown achieved by the static policy and plotted in Figure 6.8a. It is observed that best results are obtained with $V_{thr} = 0.85$. With $V_{thr} = 0.7$ performance slightly decreases. On the other hand, as the switching threshold is increased above 0.85, the performance gets worse. When V_{thr} is set to 0.95, the slowdown increases especially for small arrival rates. However, as the network load increases, relative slowdown obtained with $V_{thr} = 0.95$ gets better and at $\lambda = 0.9$ best slowdown is achieved with $V_{thr} = 0.95$. When the threshold is set to 1, the policy corresponds to the static policy and attains the worst performance as discussed before.

The results shown in Figure 6.8b suggest that the fairness is relatively insensitive to the value of V_{thr}. However, there is a sharp decrease in fairness as the threshold value gets very close to 1. Holding cost results presented in Figure 6.8c is similar to the slowdown results. For small λ, best results are obtained with $V_{thr} = 0.7$ and 0.85. It is observed that for high network load levels minimum holding cost is achieved with $V_{thr} = 0.95$.

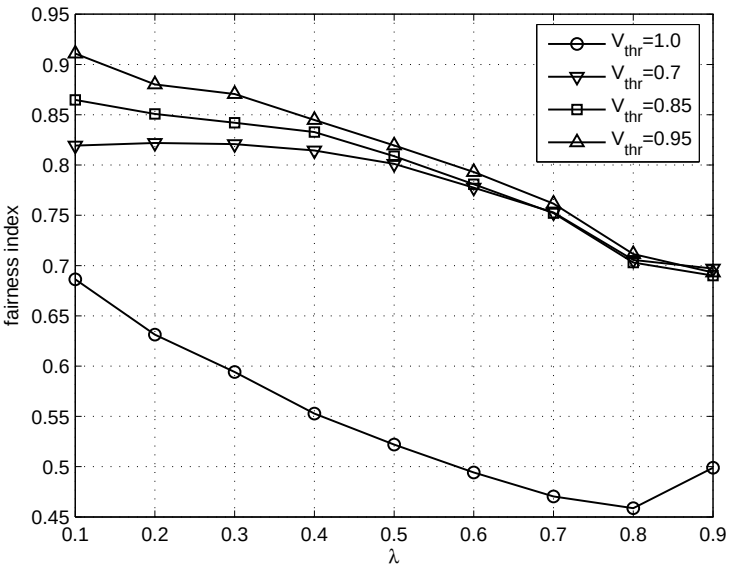
Average number of reconfigurations presented in Figure 6.8d clearly demonstrate that as the threshold is increased, rate of reconfiguration decreases gradually. And for $V_{thr} = 1$, no reconfigurations are performed as expected.

To further analyze the effect of V_{thr} on the slowdown performance of the HM3 method, slowdown results are obtained at different load levels using V_{thr} values in the range of 0 to 1. Then, the sub-optimality of HM3 with respect to the optimum slowdown achieved with NSFS method is calculated at each load level and V_{thr} value, and converted into percentages. The results are plotted as contours in Figure 6.9 where x- and y-axis correspond to the arrival rate and V_{thr}, respectively. For instance, for $\lambda = 0.7$ and $V_{thr} = 0.85$, it is read that, HM3 is 2% worse than the optimum policy in terms of slowdown. Since the results do not change for $V_{thr} < 0.7$, the figure is drawn for V_{thr} values larger than 0.7.

With the inspection of Figure 6.9, the following observations are made. The most suitable choice of V_{thr} for this network is 0.85, and the optimality gap of HM3 is below 5% for all load levels with this selection. It can also be observed that

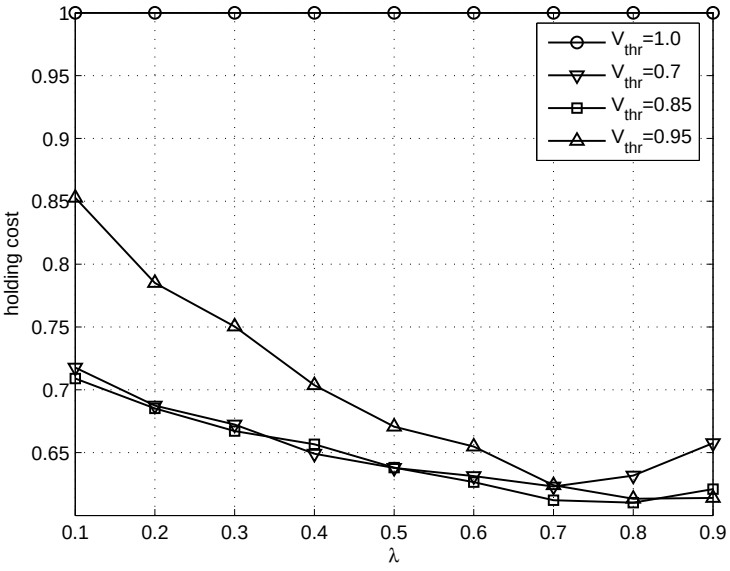


(a) Slowdown

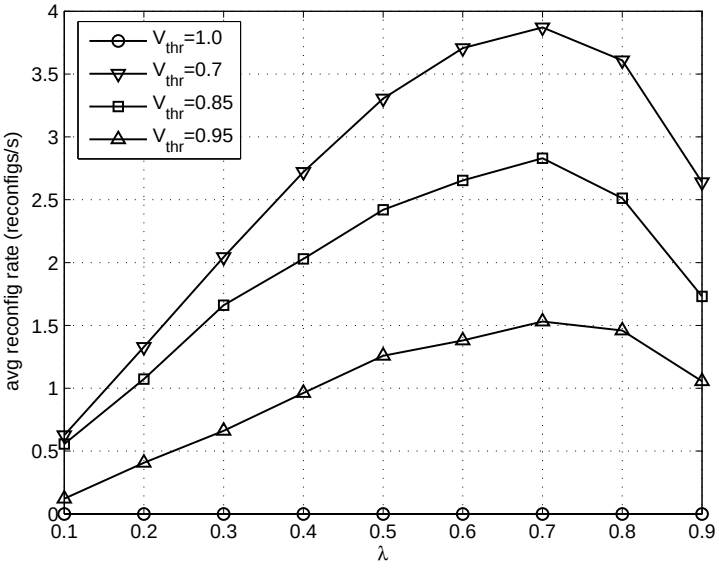


(b) Fairness

Figure 6.8: Sensitivity of HM3 performance to V_{thr} parameter.



(c) Holding Cost



(d) Average Switching Rate

Figure 6.8 (continued): Sensitivity of HM3 performance to V_{thr} parameter.

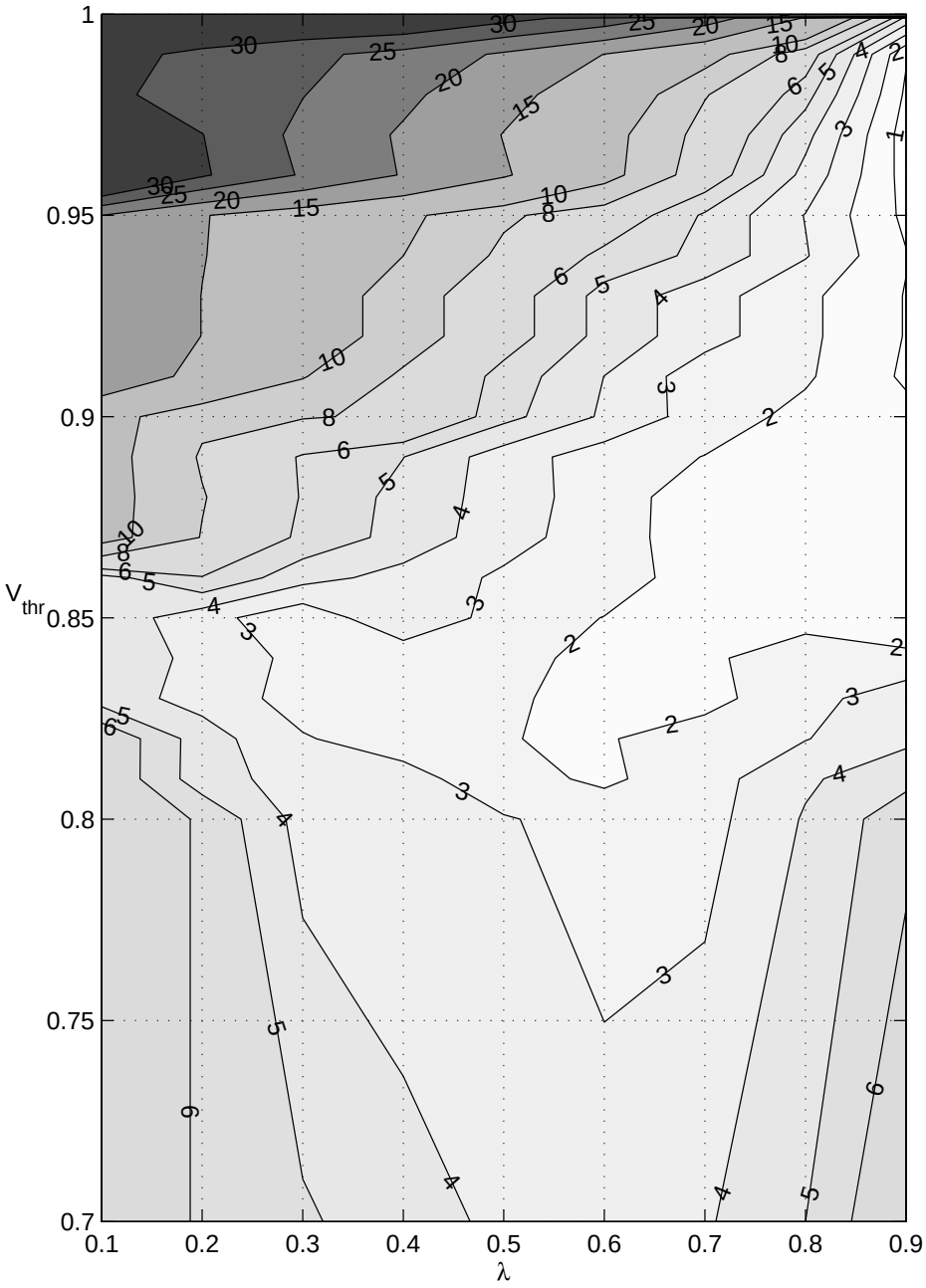


Figure 6.9: Sub-optimality of HM3 (in percentage) as a function of V_{thr} for different arrival rates.

the minimum sub-optimality achieved is 5% for $\lambda = 0.1$, whereas it is possible to decrease the sub-optimality to below 1% for $\lambda = 0.9$ using an appropriate V_{thr}. Hence, it can be concluded that HM3 gets closer to optimum policy as the arrival rate is increased. The worst results are obtained when V_{thr} is close to 1, which corresponds to the static allocation as discussed before. It is seen that the sub-optimality is not very sensitive to the threshold value used as long as it is not close to 1. It is also possible to state that as the network load increases, it may be better to use a larger V_{thr}.

6.4 Effects of Traffic and Network Parameters on the Performance of Heuristic Methods

Adaptability of a DWA method to different networks and under different traffic characteristics is an important feature, especially when the dynamics of the traffic in metro access networks are considered. In the previous sections of this chapter, the performance of the heuristic methods are evaluated as a function of flow arrival rates and it is shown that HM3 performs well for all levels of network load in terms of both throughput and fairness criteria. In this section, the sensitivities of heuristic methods to other relevant traffic and network parameters are studied. First, the effects of average flow size and channel bandwidth are considered, followed by discussions on average reconfiguration delay and total number of wavelengths in the network.

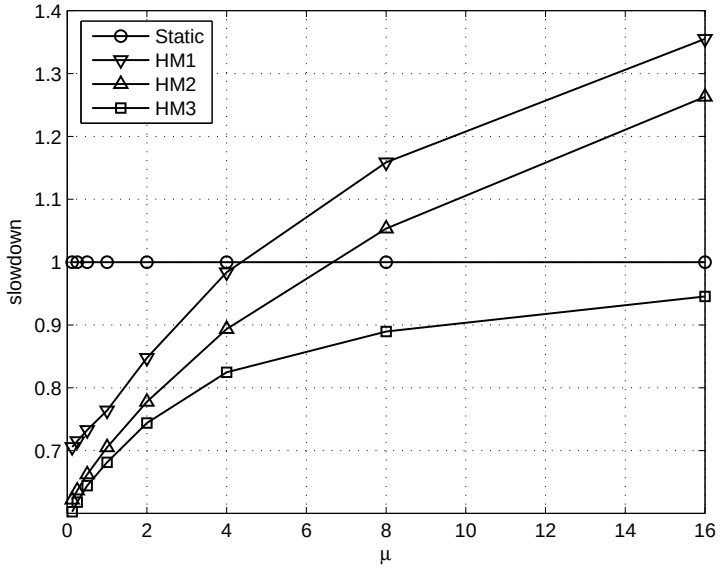
6.4.1 Average Flow Size and Channel Bandwidth

The service rates of flows at each node i by a single wavelength channel is determined by the average flow size and channel bandwidth, which can be expressed as $\mu_i = B/FS$, where B is the bandwidth of a single channel and FS is the average flow size. Without loss of generality, in the rest of this section it is assumed that $\mu_i = \mu$ for all i.

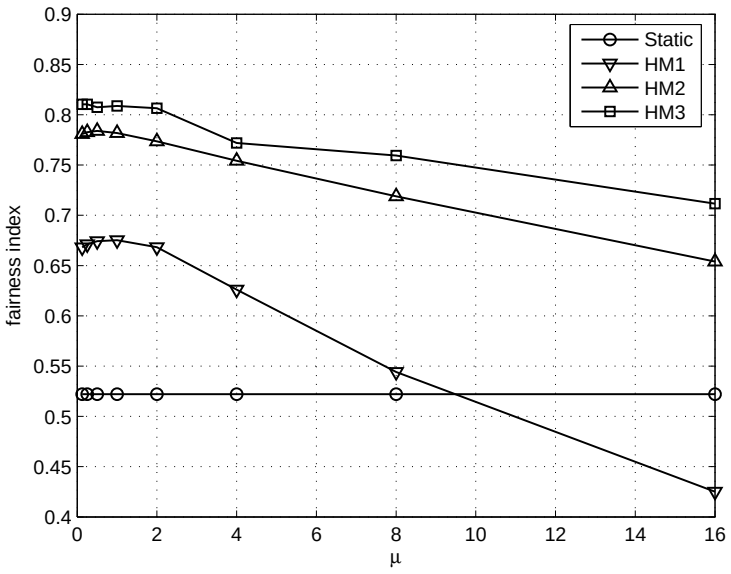
Increasing the average flow size is equivalent to decreasing the channel bandwidth, and vice versa. Average flow size determines the service rate of flows in the network and it may be argued that for a given load level, uncertainty increases with decreasing flow size (increasing arrival rate). At the extreme case, as the average flow size shrinks to zero, current state of the network carry no information about the future states, since flow service rates tend to infinity. At such a case dynamic reconfiguration turns into a vain effort. Therefore, it may be expected that the switching policy should make less and less switches and converge to the static policy as the flow sizes decrease. Same reasoning also holds for increasing channel bandwidth.

To evaluate the effects of μ on the performance of heuristic methods, simulations are performed on the 3-node test network given in Figure 4.4. The load is kept fixed at 0.5 and the average flow size and flow arrival rates are changed accordingly. The results are presented in Figure 6.10.

When slowdown and holding cost curves are inspected, it is readily seen that only HM3 successfully adapts to cases with different values of μ. For large μ, performance of HM1 and HM2 deteriorate rapidly and become even worse than the static policy. On the other hand, HM3 adjusts itself appropriately and converges to the static policy as μ is increased. This behavior is also observed from the switching rate results. It is seen that HM1 does not decrease the number of reconfigurations as μ is increased. On contrary, wavelength switching is discouraged with HM2 and HM3 and they make smaller number of reconfigurations at each value of μ. However, the switching decisions of HM3 turn out to be more efficient than HM2. In terms of fairness, HM3 again performs the best with a significant improvement over the static policy. HM2 also achieves high levels of fairness and with HM1 fairness drops below the level obtained by static allocation for $\mu > 8$.

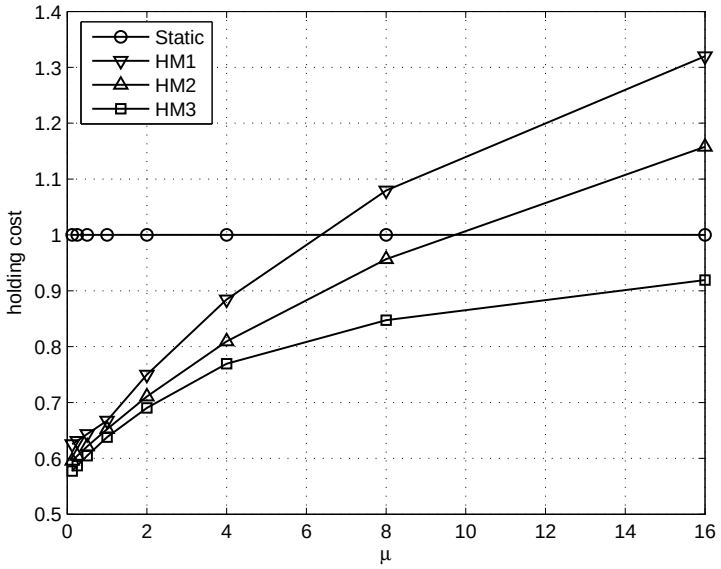


(a) Slowdown

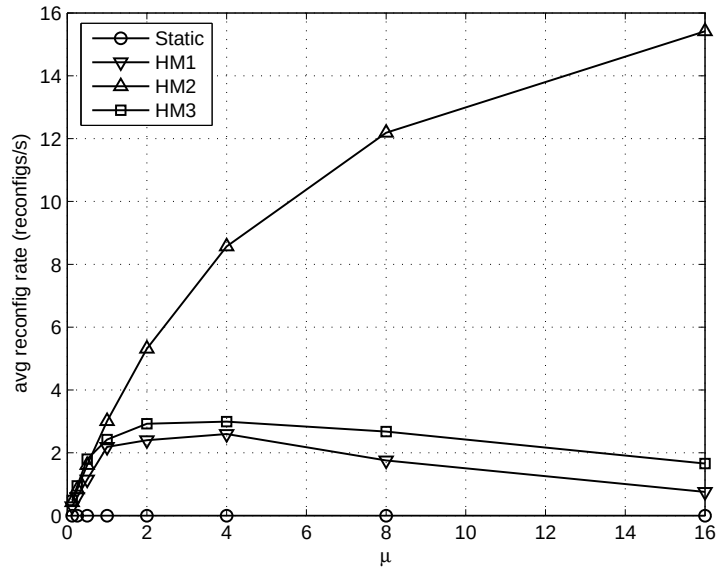


(b) Fairness

Figure 6.10: Comparison of heuristic methods for different average flow sizes.



(c) Holding Cost



(d) Average Switching Rate

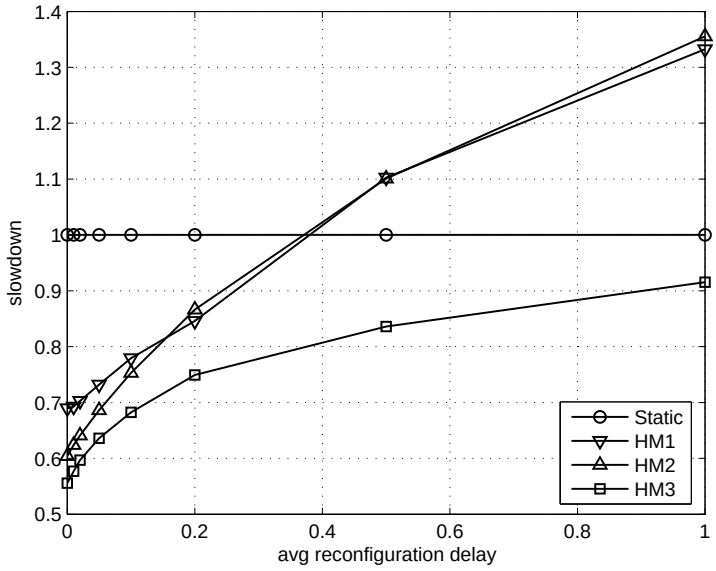
Figure 6.10 (continued): Comparison of heuristic methods for different average flow sizes.

6.4.2 Average Reconfiguration Delay

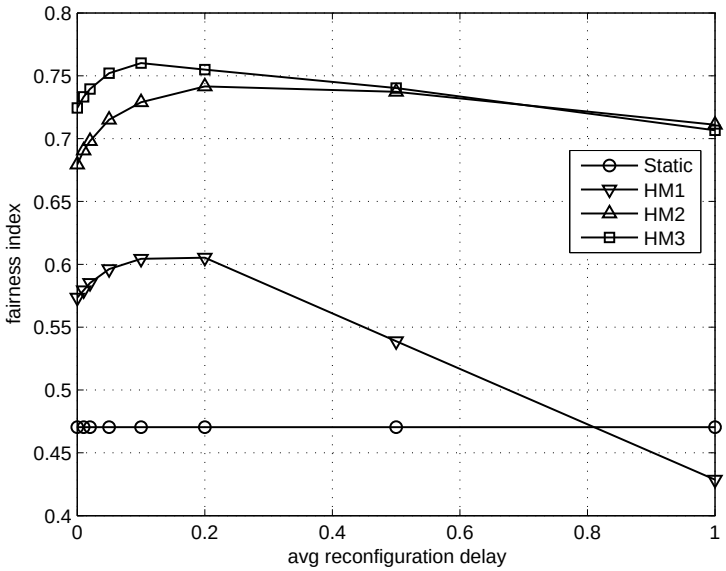
The cost of reconfiguration depends on the reconfiguration delay, during which the capacity of a whole wavelength channel is lost and as a result service rate of the network is reduced. Moreover, due to the single switch constraint, further reconfigurations are delayed during this time period. As the average reconfiguration delay is increased the cost of reconfiguration gets higher and it may be argued that the DWA method should make less number of wavelength switches in this case.

The heuristic methods are used to obtain simulation results with different values of average reconfiguration delay for the 3-node test network given in Figure 4.4. The arrival rates are 0.7, 1.4, and 2.8 flows/sec for node 1, node 2 and node 3, respectively. Service rate of a flow by a single wavelength channel, μ_i, is 1 for all nodes i. Hence, the overall network load is 0.7. Average reconfiguration delay take values beginning from 0 up to 1 seconds. Figure 6.11 plots the performance of the heuristic methods as a function of the average reconfiguration delay.

The reconfiguration rate curves show that for small values of reconfiguration delay, HM3 performs the largest number of reconfigurations. As the reconfiguration delay is increased, each policy makes less number of reconfigurations. However, the decrease rate is larger for HM1 and HM3. At reconfiguration delay of 1 seconds HM2 performs nearly maximum number of reconfigurations possible, which is equal to 1 reconfigs/sec. The slowdown and holding cost curves for HM1 and HM2 are similar. Their performance is reasonable for small reconfiguration delay and fall rapidly as the delay value is increased. At and above the point where delay is 0.5 seconds, they perform worse than the static policy. On the other hand, HM3 achieves minimum slowdown and holding cost for the whole range of reconfiguration delay, which indicates that it successfully adapts itself to rising reconfiguration costs. For reconfiguration delay of 1 seconds, it makes only 1 reconfiguration at 10 seconds on the average and still improves the slowdown and holding cost nearly 10% with respect to the static policy. The fairness performance of HM3 and HM2 are much better than HM1 and static policy. And

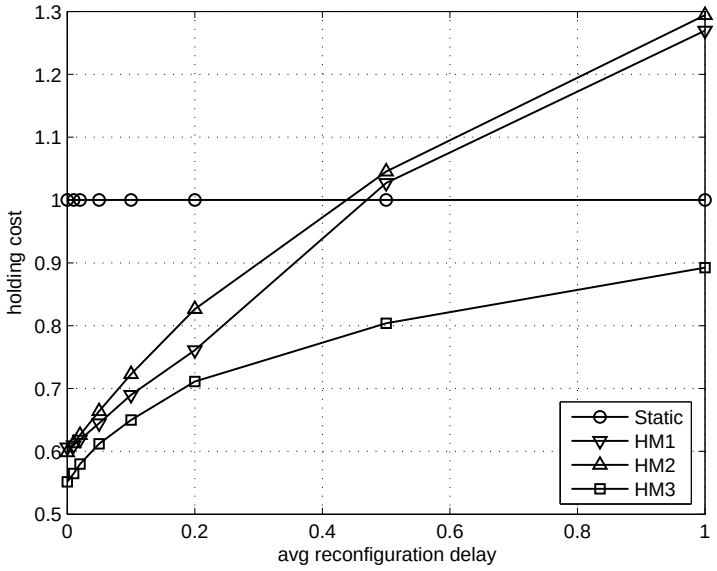


(a) Slowdown

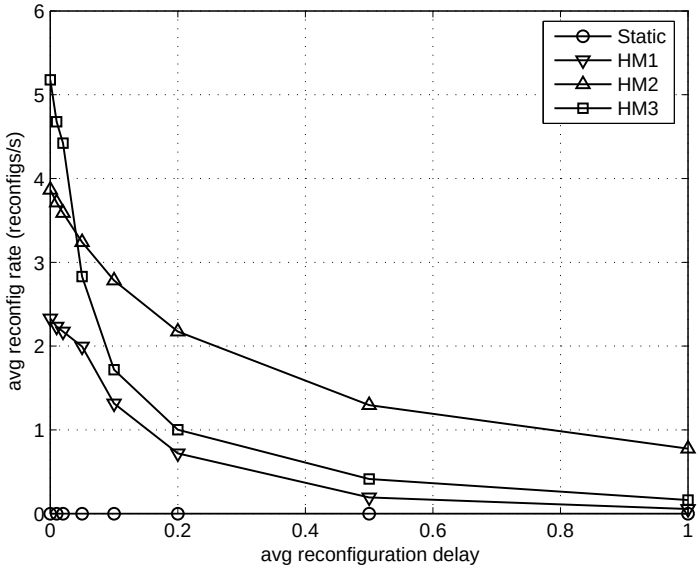


(b) Fairness

Figure 6.11: Comparison of heuristic methods for different average switching delay values.



(c) Holding Cost



(d) Average Switching Rate

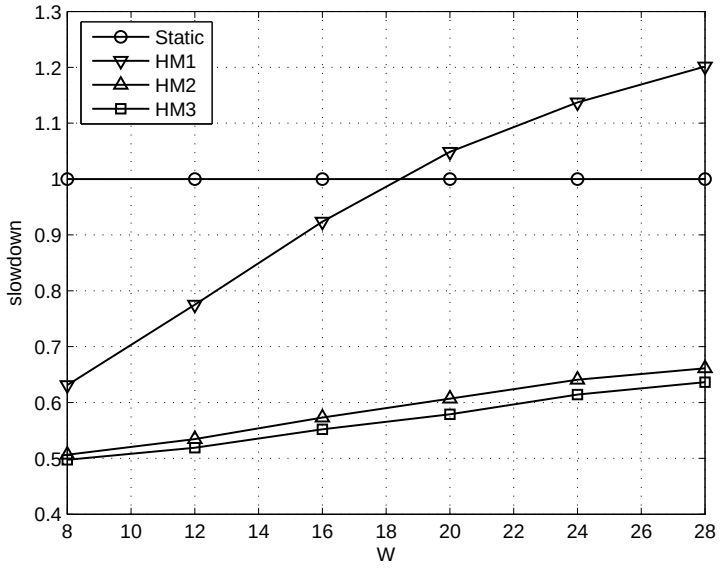
Figure 6.11 (continued): Comparison of heuristic methods for different average switching delay values.

HM1 performs worse than the static policy when the delay value is 1 sec. As a result, it can be concluded that HM3 consistently outperforms other policies in terms of any performance criteria and at all reconfiguration delay values.

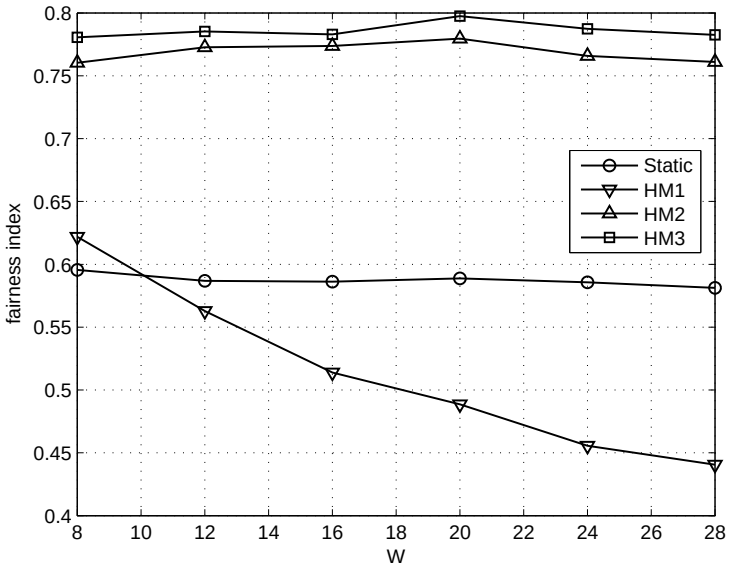
6.4.3 Total Number of Wavelengths

Finally, a 4-node test network is used to evaluate the performance of the heuristic methods as a function of the total number of wavelength channels available. The network load is kept fixed at 0.8 by scaling the arrival rates at each node with W. A uniform traffic pattern is assumed where the arrival rate at each node is equal. Hence, for the static policy the wavelengths are evenly distributed between nodes. W is changed from 8 to 28 with steps of 4 wavelengths and results are plotted in Figure 6.12.

A consistent behavior is observed for HM2 and HM3. They achieve 35-50% and 40-50% improvements in slowdown and holding cost, respectively, when compared to the static allocation. The fairness is also well above the static policy. Comparing HM2 and HM3, it can be stated that HM3 performs smaller number of reconfigurations and attains lower slowdown and holding cost for any value of W. The fairness is also better with HM3. On the other hand, HM1 shows a problematic behavior. It has the worst performance among the heuristic policies. Moreover, as the number of wavelengths is increased the performance of HM1 further decreases. Slowdown is worse than the static policy for $W > 18$. The fairness is below the static policy even for small number of wavelengths. Although holding cost is better than static allocation, the improvement vanishes as W is increased.

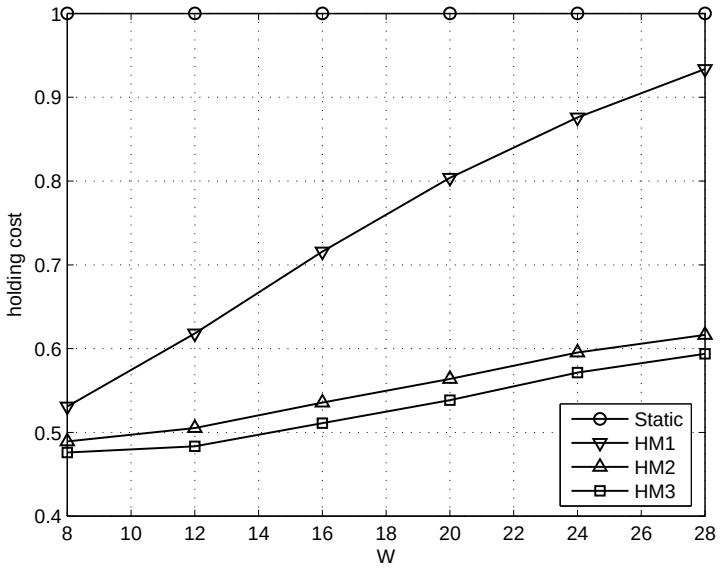


(a) Slowdown

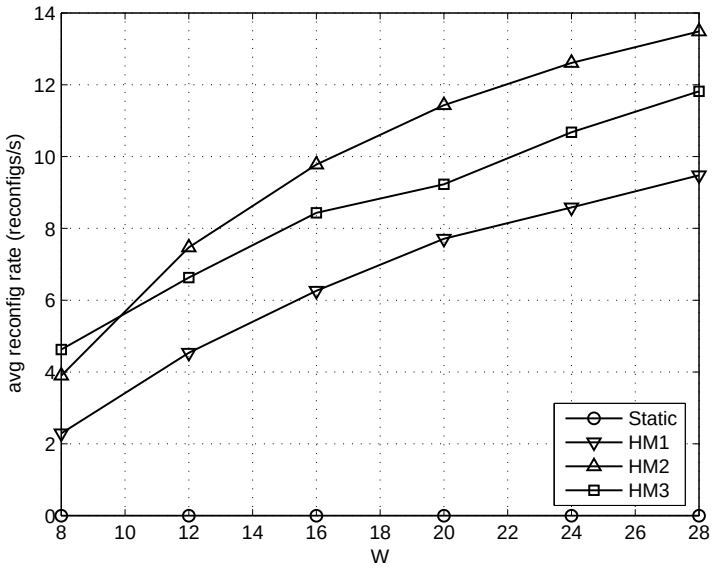


(b) Fairness

Figure 6.12: Comparison of heuristic methods for different total number of wavelengths.



(c) Holding Cost



(d) Average Switching Rate

Figure 6.12 (continued): Comparison of heuristic methods for different total number of wavelengths.

Chapter 7

Performance Bounds for DWA

In the previous chapters, DWA methods are compared with the static wavelength allocation scheme in order to determine the relative performance of each DWA policy. In this chapter, lower bounds for the average flow duration are discussed so that maximum performance gain achievable with dynamic allocation is obtained. These bounds are also used to evaluate the extent of potential performance enhancement attained by the DWA methods.

The complexities introduced by the constraints of the DWA problem make it hard to calculate strict lower bounds. However, two lower bounds are obtained in this chapter by relaxing some of these constraints. A basic lower bound is calculated by modeling the system as an M/M/1-PS queue with homogeneous transition rates. A tighter bound is developed with the introduction of the connectivity constraint and constructing an inhomogeneous Markov chain. The bounds are then demonstrated on a sample scenario, and the effects of the single switching constraint and reconfiguration delay are discussed.

7.1 Lower Bound 1 (LB1)

In order to obtain a lower bound for average flow duration, the problem is idealized as follows:

1. In the DWA problem, each node is allocated at least one wavelength to satisfy the connectivity requirement. To obtain a lower bound this constraint is relaxed, and it is allowed that a node can be assigned no wavelengths.

2. Reconfiguration delay is neglected. Hence, wavelength reconfiguration is assumed to be instantaneous.

3. In the original problem, at most one wavelength is allowed to be in the switching state. To find a lower bound, single switch constraint is also relaxed. So, the wavelength distribution can be reconfigured arbitrarily at a flow arrival or departure instant.

4. Instead of channel switching, an infinitesimal bandwidth granularity is assumed, so that a perfect load balance can be achieved.

5. For simplicity service rate of flows by a single wavelength channel at each node is assumed to be equal, i.e., $\mu_i = \mu$, $\forall i$.

With the above assumptions, the network is equivalent to a single M/M/1-PS queue where flows arrive with rate $\Lambda = \sum_{i=1}^{N} \lambda_i$ and the service rate is $M = W\mu$.

The number of customers in an M/M/1-PS queue is represented with the same birth-death process as in the M/M/1-FCFS queue [42]. Hence, the probability that there are f flows in the system, p_f, is given by

$$p_f = (1 - \rho)\rho^f,$$

where $\rho = \Lambda/M$. The expected number of flows in the queue is

$$E[f] = \frac{\rho}{1 - \rho}.$$

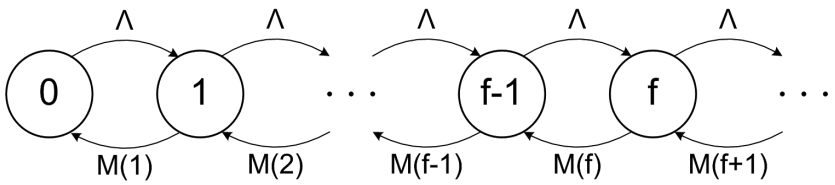


Figure 7.1: Markov chain corresponding to the number of flows in the network.

Using Little's Theorem, the expected service time (flow duration) can be found as:

$$LB1 = E[T] = \frac{1/M}{1-\rho} = \frac{1}{M-\Lambda}$$

7.2 Lower Bound 2 (LB2)

A tighter lower bound can be calculated considering that at least one wavelength should be allocated to each node. Because of this constraint, if there are no flows present at a node, then the wavelength assigned to that node will be idle and the total service rate of the network is decreased. For example, for a 3-node network with 7 wavelength channels, if there are flows only at a single node, then the total service rate will be 5μ, and if there are flows at two of the nodes, the total service rate will be 6μ. The Markov chain corresponding to this modified process is shown in Figure 7.1 . This chain is no more homogeneous since the service rate depends on the state (i.e., the number of flows at the network).

The service rate, $M(f)$, can be written as:

$$M(f) = \begin{cases} (W-N+1)\mu, & \text{with probability P(f,1)} \\ (W-n+2)\mu, & \text{with probability P(f,2)} \\ \vdots & \vdots \\ W\mu, & \text{with probability P(f,N)} \end{cases}$$

where W is the total number of wavelengths available, N is the number of nodes and $P(f,n) = \Pr\{\text{there are flows at } n \text{ distinct nodes} \mid \text{there are } f \text{ flows in the network}\}$.

Obviously, $P(f,n) = 0$ if $f < n$. For $f \geq n$, $P(f,n)$ can be calculated

recursively, assuming that flows at all nodes are served with equal rates. The following lemma gives an expression for $P(f,n)$.

Lemma 8. *If the set of all nodes $A = \{1, ..., N\}$, total number of flows is f and probability that a flow belongs to node i is r_i, then the probability that all the flows are distributed among any set of n nodes is given by*

$$P(f,n) = \sum_{z \in \mathcal{S}_n^A} \left(\sum_{j \in z} r_j \right)^f - \sum_{k=1}^{n-1} P(f,k) \frac{\binom{n}{k}\binom{N}{n}}{\binom{N}{k}}$$

where $r_i = \lambda_i / \Lambda$ is the probability that a flow in the system belongs to node i and $\mathcal{S}_n^A$ is the set of all subsets of the set $A = \{1, 2, \ldots, N\}$ with n elements.

Proof. Let $\mathcal{S}_n^A$ denote the set of all subsets of A with n elements. Define

$$p(z,f,n) = \Pr\left(\mathcal{C}\left(i \in z : f_i > 0\right) = n \mid \sum_{i \in A} f_i = f \right)$$

where $\mathcal{C}(x)$ denotes the cardinality of set x. So, $p(z,f,n)$ is the probability that n of the nodes in the set z have a total of f flows and the rest of the nodes have no flows.

$P(f,n)$ can be written as

$$P(f,n) = p(A,f,n) = \sum_{z \in \mathcal{S}_n^A} p(z,f,n)$$

and $p(A,f,n)$ can be calculated as follows

$$
\begin{aligned}
p(A,f,n) &= \sum_{z\in\mathcal{S}_n^A}\left(\Pr(f_i=0,\forall i\notin z)-\sum_{k=1}^{n-1}\sum_{z'\in\mathcal{S}_k^s}p(z',f,k)\right)\\
&= \sum_{z\in\mathcal{S}_n^A}\left(\sum_{i\in z}r_i\right)^f-\sum_{z\in\mathcal{S}_n^A}\sum_{k=1}^{n-1}\sum_{z'\in\mathcal{S}_k^s}p(z',f,k)\\
&= \sum_{z\in\mathcal{S}_n^A}\left(\sum_{i\in z}r_i\right)^f-\sum_{k=1}^{n-1}\sum_{z\in\mathcal{S}_n^A}\sum_{z'\in\mathcal{S}_k^s}p(z',f,k)\\
&= \sum_{z\in\mathcal{S}_n^A}\left(\sum_{i\in z}r_i\right)^f-\sum_{k=1}^{n-1}\frac{\binom{n}{k}\binom{N}{n}}{\binom{N}{k}}\sum_{z\in\mathcal{S}_k^A}p(z,f,k)\\
&= \sum_{z\in\mathcal{S}_n^A}\left(\sum_{i\in z}r_i\right)^f-\sum_{k=1}^{n-1}\frac{\binom{n}{k}\binom{N}{n}}{\binom{N}{k}}P(f,k)
\end{aligned}
$$

□

For illustration, $P(f,n)$ values are calculated for a network with 3 nodes and 7 wavelengths. Arrival rates are 0.5, 1, and 2 for node 1, node 2, and node 3, respectively. In Figure 7.2, the results are plotted as a function of total number of flows in the network. When $f=1$, it is certain that only one of the nodes has a flow and the other two nodes are empty, that is $P(1,1)=1$, $P(1,2)=0$ and $P(1,3)=0$. Similarly, $P(2,3)=0$. As the number of flows in the network increases $P(f,1)$ and $P(f,2)$ converges to 0. As f goes to infinity, the probability of having at least one flow at each node converges to 1.

After having calculated the $P(f,n)$ values as discussed above, the Markov chain shown in Figure 7.1 is solved numerically. For this purpose, the chain is truncated at a large enough f value, and the invariant distribution, π, for the resulting finite chain can be found using one of the several numerical solution techniques to solve $\pi Q=0$, where Q is the infinitesimal generator. Using the invariant distribution, expected number of flows in the system is calculated as

$$E[f]=\sum_i i\pi_i \tag{7.1}$$

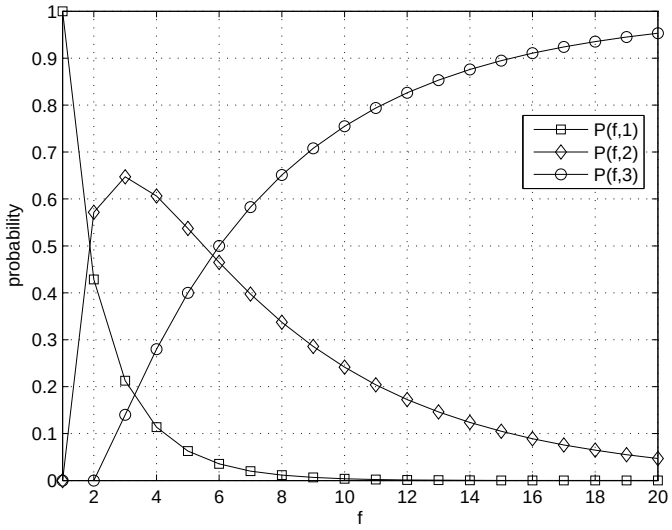


Figure 7.2: $P(f, n)$ for $N = 3$, $W = 7$, $\lambda = (0.5, 1.0, 2.0)$.

The expected average flow duration can be found once more using Little's Theorem, yielding the value of the lower bound, LB2, as

$$LB2 = E[T] = E[f]/\Lambda.$$

In the following section, lower bounds are computed on a sample network and compared with the results of optimum and heuristic DWA policies.

7.3 Numerical Results

Lower bounds developed in this chapter are demonstrated on the 3-node test network given in Figure 4.4, with $\mu_i = 1$ for all nodes i and $(1/\sigma) = 50$ms. For this scenario, optimum and heuristic policies have already been obtained in Chapter 4 and Chapter 6, respectively. First, LB1 and LB2 are compared with each other and with the static policy. Next, the effects of the single switching constraint and reconfiguration delay on the flow completion times are discussed. For this aim, LB2 is compared with the optimum FS policies for the case with zero reconfiguration delay and single switching constraint relaxed. Since, FS policy is

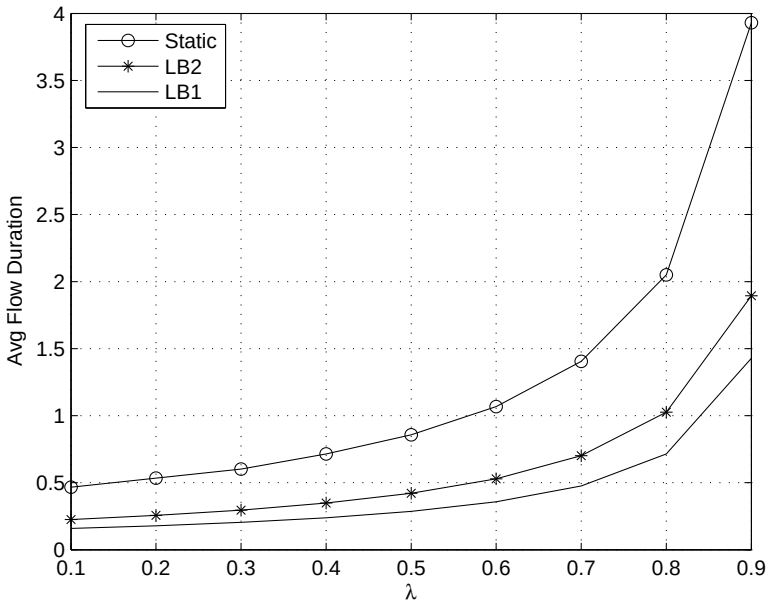


Figure 7.3: Comparison of LB1, LB2 and static policy.

known to minimize the average flow completion time, this comparison highlights the effects of each factor. Finally, the performance of NSFS policy and HM3 are compared with LB2.

7.3.1 Comparison of Lower Bounds with the Static Policy

To compare LB1 and LB2 with each other and with the static policy, average flow completion times are obtained using the 3-node test network given in Figure 4.4. Figure 7.3 plots the values of LB1 and LB2, along with the results of the static policy, as a function of arrival rate λ.

The difference between the curves LB1 and LB2 is due to the connectivity constraint, which states that at least one wavelength should be allocated to each node at any time. The wavelengths allocated to a node are idle unless there is at least one flow at that node. This causes a loss in the total network capacity and an increase in average flow completion time. When λ is small, the probability of having no flows at a node is higher and the capacity lost due to the connectivity

constraint is larger. As the arrival rate is increased, it becomes less likely to have no flows at a node and the expected time that wavelengths remain idle gets smaller. However, the dependence of the average flow duration on the network load is not linear. As the load approaches 1, the slope tends to infinity. Due to this fact, the effect of a given amount of capacity loss on the average flow duration is more pronounced as the arrival rates are increased. These two effects nearly cancel each other and it is observed that LB2 is 32-48% larger than LB1 for all λ.

Comparison of LB2 with the results of static policy shows that DWA can potentially improve the flow completion times by 50% at all λ, in the absence of the single wavelength switching constraint and reconfiguration delays.

7.3.2 Effects of Single Wavelength Switching Constraint and Reconfiguration Delay

The effects of the single wavelength switching constraint and reconfiguration delays are demonstrated in Figure 7.4, where the flow completion times are normalized with respect to LB2. FS cost function is used to obtain optimum policies since it minimizes the average flow duration as discussed in Section 4.4. FS_0^{MS} corresponds to the FS policy obtained for the case where more than one wavelength can be switched at each decision instant without reconfiguration delays. It is verified that the flow completion times achieved with FS_0^{MS} is very close to LB2. FS_0 is the policy obtained using the cost function FS with single switching constraint and without reconfiguration delays. Therefore, the area between the curves FS_0^{MS} and FS_0 may be attributed to the cost incurred due to the single wavelength switching constraint. The difference between these two policies is larger for small λ and vanishes as the arrival rates are increased.

This behavior can be explained as follows. When λ is small, the expected number of flows in the network is also small and the arrival or departure of a single flow may require more than one wavelength switchings to minimize the average flow duration. For instance, suppose that the number of flows and wavelengths at

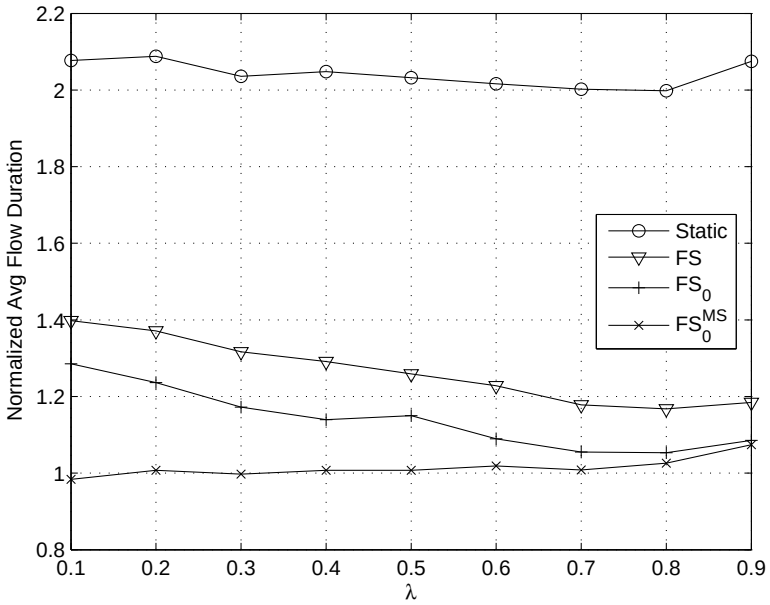


Figure 7.4: Effects of single wavelength switching constraint and reconfiguration delay on the performance of DWA methods.

each node are given by $f = (1, 2, 1)$ and $w = (4, 2, 1)$, respectively. The departure of the flow at node 1 requires the switching of 3 wavelengths from node 1 in order to minimize the capacity loss. However, for large λ, the number of flows at each node is probably higher and single wavelength switching proves to be sufficient. FS_0 results in 30% higher flow completion times with respect to FS_0^{MS} at $\lambda = 0.1$ and the difference is just 1% for $\lambda = 0.9$. Finally, FS policy is used to obtain optimum results in the existence of single wavelength switching constraint and non-zero reconfiguration delays. The difference between the FS and FS_0 curves is due to the delays associated with switching actions. Comparing FS and LB2 curves, it can be concluded that LB2 is only 20-40% lower than the minimum average flow duration achievable in this case.

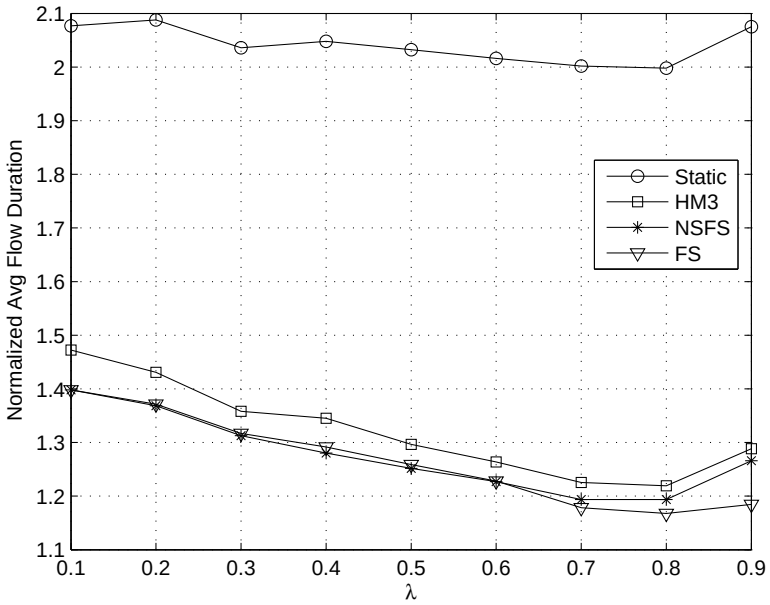


Figure 7.5: Comparison of NSFS and HM3 with LB2.

7.3.3 Comparison of LB2 with Optimum Policies and HM3

The results obtained with NSFS and HM3 policies are compared with the FS policy and LB2 in Figure 7.5. It can be observed that the optimum policies NSFS and FS are very close each other. The difference between these two curves correspond to the cost paid by the NSFS policy for improving the slowdown and fairness performance. It is seen that up to very high arrival rates, FS and NSFS performs nearly same in terms of average flow duration. For large λ, NSFS performs larger number of reconfigurations to balance the load in the network which results in a decrease in the average flow throughput. The results obtained with HM3 is also plotted for comparison purposed. As discussed in Chapter 6, the area between the curves NSFS and HM3 corresponds to the suboptimality of the heuristic and it narrows down as the arrival rate is increased.

Chapter 8

Topics for Future Research

In this chapter, possible extensions of the work discussed in this thesis and directions for future research are highlighted. Some preliminary results are also presented where applicable.

In HM3 algorithm, the value of the V_{thr} parameter is fixed. It is clear, that V_{thr} may have important effects on the performance of HM3, e.g., the optimum value of V_{thr} changes as λ varies as it was shown in Section 6.3. Hence, a modification of HM3 algorithm such that V_{thr} is adaptively set by the algorithm is identified in Section 8.1 as a possible future research area.

In the DWA framework developed in Chapter 3, it is assumed that the flows are elastic and perfect process sharing is achieved such that the capacity allocated to a node is efficiently utilized in a fair manner by the flows at that node. In Section 8.2, the applicability of this assumption to real life TCP (Transmission Control Protocol) flows are questioned and possible future extensions are discussed.

8.1 Adaptive Tuning of the V_{thr} Parameter

HM3 algorithm uses the V_{thr} parameter to decide whether the long term expected rewards exceed the reconfiguration costs. An action is initiated if the value of that action exceeds V_{thr}. In the HM3 method, a constant V_{thr} value is used and it is set to 0.85 to obtain numerical results. In Section 6.3, the performance of HM3 is analyzed for different values of the V_{thr} parameter. It is discussed that as the value of V_{thr} is increased, HM3 makes less reconfigurations and when $V_{thr} = 1$, HM3 is equivalent to the static policy. V_{thr} can be set to any value in the interval [0,1] and the optimum value of V_{thr} may depend on network and traffic parameters.

In Figure 6.9, the slowdown performance of the HM3 algorithm at different network load levels is analyzed as a function of the V_{thr} parameter. For this test network, it is observed that for small values of arrival rate, optimum V_{thr} value is around 0.84. However, for $\lambda > 0.7$ the optimum threshold begins to increase with increasing arrival rate. At $\lambda = 0.9$, best results are achieved when the V_{thr} is set to 0.95.

A future direction for research is the determination of best V_{thr} for the network and traffic under consideration. With this information HM3 method can be extended such that the threshold value is dynamically adapted to the changing traffic parameters, such as the network load.

8.2 TCP Behavior and Its Effects on DWA

TCP is commonly used to adjust the transmission rates of sources and control the congestion in the network. It is a distributed algorithm which is used to share the network resources in an efficient and fair manner. Since the availability of resources and the set of competing users vary unpredictably over time, TCP necessarily uses a feedback control, where traffic sources dynamically adapt their rates in response to the congestion on their paths. As a result, TCP has a complex

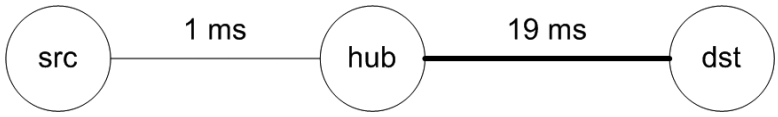


Figure 8.1: Test network used to demonstrate the effects of TCP.

behavior which is not easy to analyze.

The congestion control algorithm in the current TCP was developed in 1988 [67], and is referred as Reno. It has gone through several enhancements since its introduction and performed remarkably well as the Internet is scaled up by six orders of magnitude in size, speed, load, and connectivity. However, it is a well known fact that as the delay-bandwidth product of the transmission path increases, the performance of TCP Reno gets poor and it becomes unstable [68].

Recently, several new TCP versions (e.g., FAST TCP [69], TCP Vegas [70], High Speed TCP (HSTCP) [71], Scalable TCP (STCP) [72]) have been proposed to improve the performance of TCP for connections with large delay-bandwidth product. Like TCP Reno, HSTCP and STCP relies on packet loss events as congestions indications. TCP Vegas and FAST TCP, on the other hand, are based on the idea of measuring congestion by estimating the queueing delay.

To demonstrate the effects of TCP flow control and congestion avoidance on the performance of DWA, simple simulations are performed with ns-2 [73], using the FAST TCP module [74]. The test network is composed of three nodes as shown in Figure 8.1. The src-hub link represents the metro access network connection, and the hub-dst link corresponds to a backbone path. The src-hub and hub-dst links have propagation delays of 1 ms and 19 ms, respectively. Hence, the round trip delay (RTT) is 40 ms for this connection. The capacity of the hub-dst link is also taken much larger than the src-hub link. So, the src-hub link is the bottleneck for this path. The buffer size at the src node is taken to be 5000 packets, which corresponds to the bandwidth-delay product of the path when the capacity of the src-hub link is 1 Gbps.

Three flows from src node to dst node are considered, where the first flow is always active and other two flows are arriving and departing periodically, as

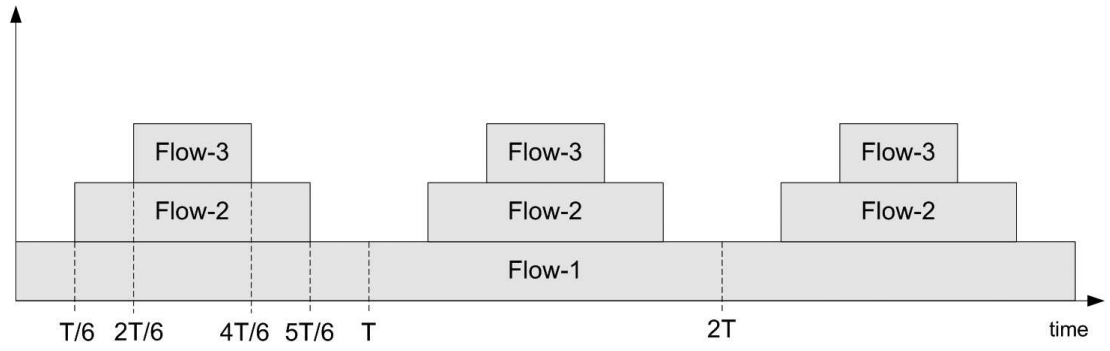


Figure 8.2: Periodic change of the number of flows from source node to destination node.

depicted in Figure 8.2, where T is the period of the traffic pattern. The average number of flows in the network is 2.

It is assumed that there are 3 wavelength channels, each with bandwidth $B = 0.5$ Gbps, and the behavior of FAST TCP is observed under two different wavelength allocation schemes. In the first case, two channels are allocated to the link between the source and hub node. In the second case, the bandwidth is changed dynamically, so that the number of channels is equal to the number of active flows at any time. Note that the average capacity allocated to the src-dst link is 2 channels in both cases. Figure 8.3 plots the throughput achieved by each flow when $T = 60$ s. The x-axis is the time, and y-axis corresponds to the bandwidth in units of single channel capacity, B. It is observed that FAST TCP efficiently utilizes the available bandwidth of the link. However, the available bandwidth is not evenly shared among the active flows. Inspecting the time period [10, 20] s, it is seen that it took 3 s before flow-2 attains a stable rate. Moreover, the bandwidth is not fairly divided between the two flows even in the steady-state. A similar behavior is observed in the time interval of [20, 30] s, where three flows are concurrently active. The departure of flows at time 40 and 50 results in some idle capacity, which is rapidly filled by the remaining flow(s). It is also observed that after the departure of flow-3, the capacity is evenly shared by flow-1 and flow-2.

The corresponding results for the dynamic bandwidth allocation case are illustrated in Figure 8.4. It is observed that with TCP FAST, flows successfully adapt

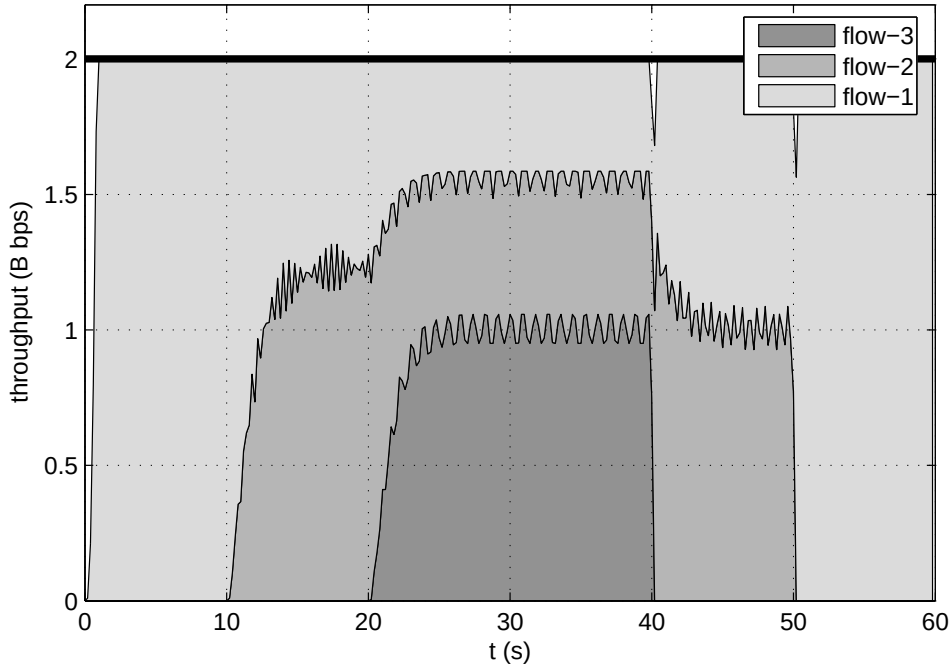


Figure 8.3: Throughput for static bandwidth allocation with $T = 60$ s.

their transmission rates so that the available capacity is utilized efficiently. As in the static case, it takes some time for the newly arriving flows to reach a stable throughput level. However, comparing with Figure 8.3, it is easily observed that after the flow rates are settled, the available capacity is fairly shared between the flows. Moreover, with the departure of flow-3 at time 40, the bandwidths utilized by flow-1 and flow-2 are rapidly adjusted to equal values.

The simulations are repeated with T is set to 6 s, and the results for the static allocation are plotted in Figure 8.5. Inspection of the figure shows that short flows are terminated before they reach a stable throughput level and once again it is observed that the available capacity is not evenly shared between active flows.

The results for the dynamic capacity allocation scheme is shown in Figure 8.6, where once again it is observed that the duration of flow-3 is not enough to achieve a stable rate. Flow-1 uses nearly a fair share of capacity approximately beginning 1.5 s after the arrival of flow-2. However, flow-2 gets a rate which is 50% higher than its fair share. It is also observed that the time required to achieve full

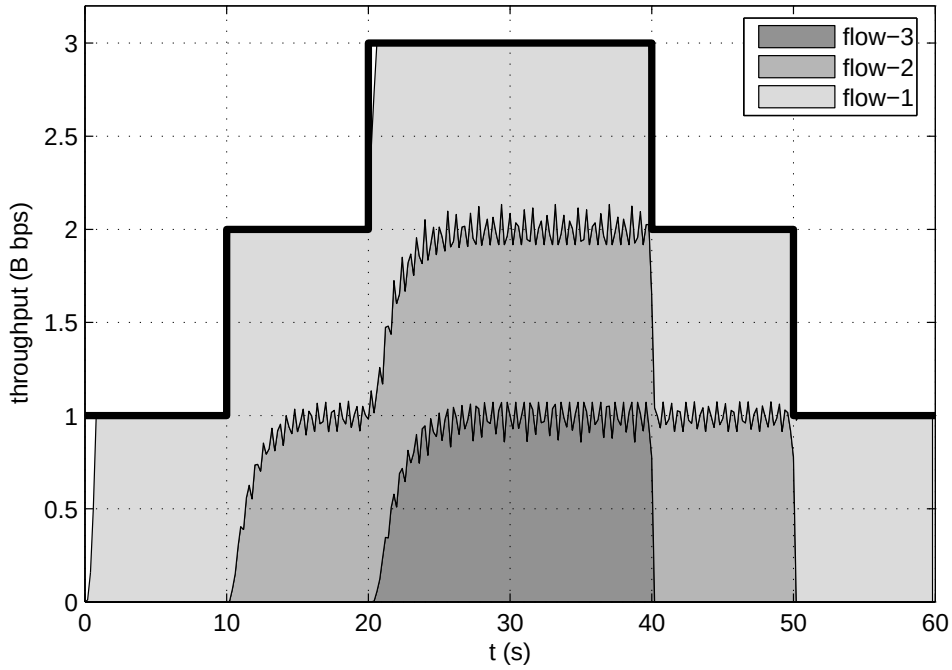


Figure 8.4: Throughput for dynamic bandwidth allocation with $T = 60$ s.

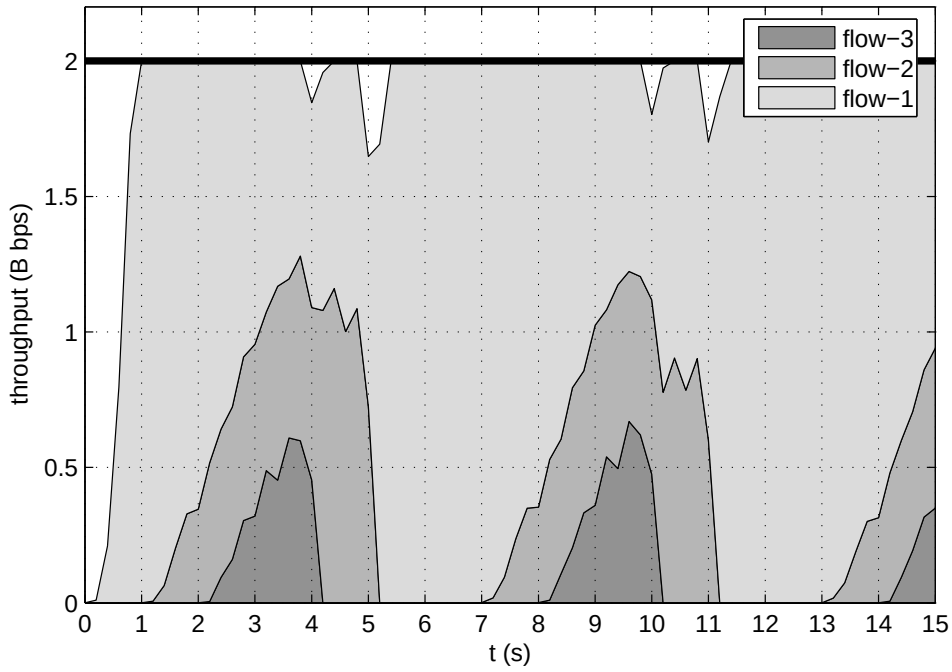


Figure 8.5: Throughput for static bandwidth allocation with $T = 6$ s.

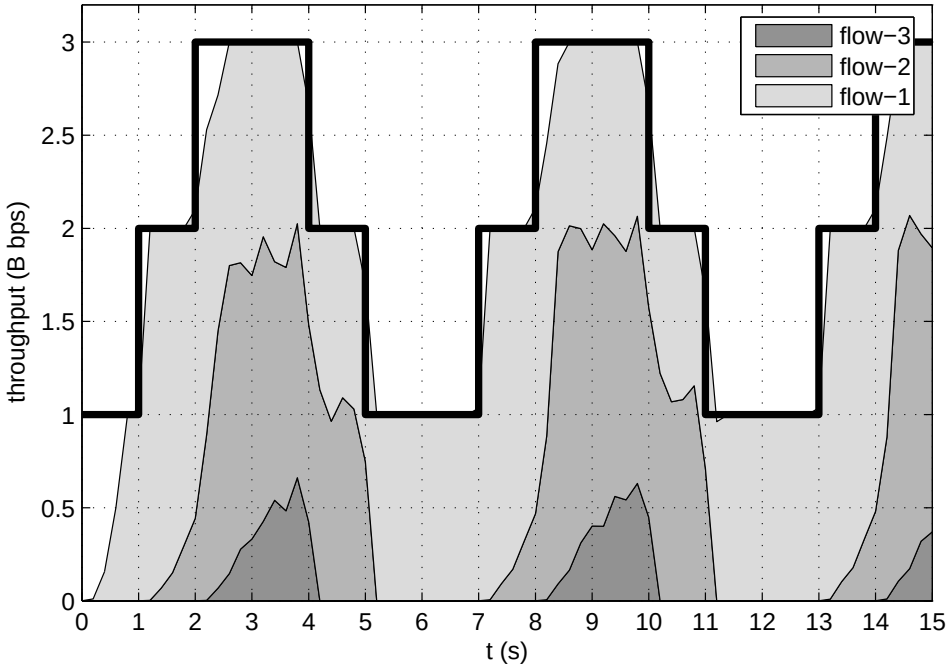


Figure 8.6: Throughput for dynamic bandwidth allocation with $T = 6$ s.

utilization after increasing the bandwidth may not be negligible when the flow durations are small.

Under the light of the above observations, the following conclusions can be drawn regarding the validity of the processor sharing model assumed in the DWA framework for the flow rates. When the flow durations are large, it can be safely assumed that the available capacity is fully utilized by the flows, and the capacity lost following a bandwidth increase is negligible. The assumption of fair capacity sharing between flows are clearly violated for both the static and dynamic wavelength allocation cases. With short flow durations, fairness cannot be achieved since the flows do not have enough time to increase their rates sufficiently. Moreover, the period of time during which the flows adapt their rates to a bandwidth increase presents a bandwidth penalty, since the available bandwidth cannot be utilized fully in this period.

To sum up, efficient and fair capacity utilization assumptions hold when the flow durations are large and the results obtained in this thesis may be valid

for such cases. However, as the flow durations get smaller, FAST TCP cannot guarantee the validity of these assumptions. Therefore, a future topic of research may be to extend the DWA methods so that the issues related to TCP control algorithm for short flows are taken into account. For instance, extension of the reconfiguration delay by a suitable amount may be thought as a possible simple solution for modeling the unutilized bandwidth following reconfiguration actions.

Chapter 9

Conclusions

With the introduction of modern access technologies based on optics and wireless, end user data rates have increased orders of magnitude. This increase has put a heavy burden on the metro access networks which had been designed and optimized for circuit switched connections. Increasing the capacity alone do not provide a long term solution. Designing wrapper protocols brings in inefficiencies and increases the complexity and cost of the system. As a result, an architectural change seems to be inevitable for metro access networks.

The prospective architecture has to meet the requirements peculiar to the metro access networks. Cost is the driving factor because a small area is served by each metro access network and cost has to be shared by a limited number of users. Protocol transparency is a desired feature due to the multitude of protocols used in distribution networks. Since each node of a metro network serves a district of a town, large traffic deviations at different time scales are expected. This may be due to factors such as the mobility of population between residential and industrial areas. Mobile wireless networking also contributes to this variability. A static design based on peak rates is inefficient in terms of capacity usage and cost. Therefore, traffic adaptability is a very important advantage for the future metro access network.

A promising proposal for the future metro access network architecture is the

wavelength routed IP/WDM ring network where lightpaths are established from (to) access nodes to (from) the hub node to forward the IP traffic over WDM. The lack of an intelligent networking layer provides simplicity and transparency. With the use of tunable transmitters, it is also possible to dynamically change the configuration of wavelengths to follow traffic fluctuations. Naturally, the reconfiguration policy has a direct impact on the efficiency and proper functioning of the network.

In this thesis, we investigated dynamic reconfiguration policies for IP/WDM metro access networks. As a first step, we constructed a proper framework for the definition of the DWA problem. With reasonable assumptions, the problem was modeled as a continuous time MDP. We obtained the optimum DWA policy by solving the MDP. The performance of the policy certainly depends on the cost function used in the optimization process. We proposed a novel cost function (NSFS) to improve both slowdown and fairness performance of the network. We showed that the resulting policy performs wavelength switches in order to improve the load balance in the network as anticipated. NSFS was also compared with other cost functions derived from the literature (FS and NFS) on a 3-node test network. We demonstrated that NSFS achieves consistently better performance in terms of both slowdown and fairness.

The MDP approach is useful to understand the characteristics of the optimum policy and obtain performance bounds but it is limited to small networks due to exponential growth of the state space. Therefore, we proposed a heuristic method (HM3). It is based on the cost function NSFS and first passage probabilities. The resulting policy was similar to the optimum policy. For a 3-node test network the optimality gap was found to be below 5% and it decreased further with the increasing network load. We also compared HM3 with other heuristics (HM1 and HM2) available in the literature. It was demonstrated that HM3 has the best performance in terms of both slowdown and fairness. For moderate and high network loads the improvement obtained in slowdown with HM3 is 37% compared to the static policy and about 10% compared to the other heuristics. Best fairness values were also obtained with HM3.

To evaluate the performance of the heuristic methods under non-stationary traffic, we used a 5-node test network and changed the average flow arrival rates with time. In this case, HM3 succeeded to obtain the best slowdown and fairness performance by making the minimum number of reconfigurations. The slowdown obtained with HM3 was 53% better than the static policy, 35% better than HM1 and 9% better than HM2. Again a significant improvement in fairness was observed with respect to other policies.

We also investigated the effects of several network and traffic parameters on the performance of heuristic policies. First, the average flow size was considered, and it was observed that HM3 is significantly better than the static policy although the improvement decreased with the decreasing average flow duration as expected. On the other hand, HM1 and HM2 performed worse than the static policy as the average flow length was decreased. Best fairness results were again obtained with the HM3 policy. Next, we discussed the effects of average reconfiguration delay. As the average reconfiguration delay and hence the reconfiguration cost was increased, HM3 successfully reduced the reconfiguration rate and always achieved better results than static policy. On the other hand, HM1 and HM2 again performed worse than the static policy for large values of average reconfiguration delay. It can be argued that unlike HM1 and HM2, HM3 consistently attains the best performance for the whole range of the average reconfiguration delay and average flow size values. Finally, we varied the total number of wavelengths to find out its effect on the performance. For this case also, it was observed that HM3 attains better results in terms of both slowdown and fairness measures.

In order to reveal the potential benefits of DWA, we investigated theoretical bounds on the performance improvement that can be achieved. The network was simplified to a single M/M/1-PS queue. The first lower bound was obtained assuming homogeneous departure rates. With the inclusion of the connectivity constraint, a time inhomogeneous Markov chain was considered and we obtained a tighter bound. Using numerical results obtained with a sample network we demonstrated that the lower bound was on the average 25% smaller than the optimum results.

Finally, we identified areas for further research. One possible extension of the current work may be the modification of the HM3 method such that V_{thr} is adaptively set by the algorithm. The potential benefits for this modification was demonstrated on a sample network. Another direction for research may be on incorporating the effects of TCP flow control which may exhibit non-ideal behavior regarding the fairness and efficient utilization of the capacity.

Bibliography

[1] A. A. M. Saleh and J. M. Simmons, "Evolution toward the next-generation core optical network," *J. Lightw. Technol.*, vol. 24, no. 9, pp. 3303–3321, May 2006.

[2] T. Koonen, "Fiber to the home/fiber to the premises: What, where, and when?" *Proc. IEEE*, vol. 94, no. 5, pp. 911–934, May 2006.

[3] D. Cavendish, "Evolution of optical transport technologies: From SONET/SDH to WDM," *IEEE Commun. Mag.*, vol. 38, no. 6, pp. 164–172, Jun. 2000.

[4] N. Ghani, "Regional-metro optical networks," in *Emerging Optical Network Technologies: Architectures, Protocols and Performance*, K. Sivalingam and S. Subramaniam, Eds. Springer, ch. 4.

[5] D. Cavendish, K. Murakami, S.-H. Yun, O. Matsuda, and M. Nishihara, "New transport services for next-generation SONET/SDH systems," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 80–87, May 2002.

[6] A. A. M. Saleh and J. M. Simmons, "Architectural principles of optical regional and metropolitan access networks," *J. Lightw. Technol.*, vol. 17, no. 2, pp. 2431–2448, Dec. 1999.

[7] G. Kramer and G. Pesavento, "Ethernet passive optical network (EPON): Building a next-generation optical access network," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 66–73, Feb. 2002.

[8] J. M. Simmons and A. A. M. Saleh, "Optical regional access networks (ORAN) project," in *Proc. Optical Fiber Communication Conference (OFC'99)*, San Diego, CA, 1999, pp. 178–180.

[9] A. A. M. Saleh and J. M. Simmons, "Architectural principles of optical regional and metropolitan access networks," *J. Lightw. Technol.*, vol. 17, no. 2, pp. 2431–2448, Dec. 1999.

[10] M. Kuznetsov, N. Froberg, S. Henion, H. Rao, J. Korn, K. Rauschenbach, E. Modiano, and V. Chan, "A next-generation optical regional access network," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 66–72, Jan. 2000.

[11] N. M. Froberg, S. R. Henion, H. G. Rao, B. K. Hazzard, S. Parikh, B. R. Romkey, and M. Kuznetsov, "The NGI ONRAMP test bed: Reconfigurable WDM technology for next generation regional access networks," *J. Lightw. Technol.*, vol. 18, no. 12, pp. 1697–1708, Dec. 2000.

[12] I. M. White, M. S. Rogge, K. Shrikhande, and L. G. Kazovski, "A summary of the HORNET project: A next-generation metropolitan area network," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 9, pp. 1478–1494, Nov. 2003.

[13] A. Carena, V. Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione, and P.Poggiolini, "RingO: An experimental WDM optical packet network for metro applications," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1561–1571, Oct. 2004.

[14] L. Dittmann, C. Develder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Sole-Pareta, W.Cerroni, N. Leligou, L. Dembeck, B. Mortensen, M. Pickavet, N. L. Sauze, M. Mahony, B. Berde, and G. Eilenberger, "The european IST project DAVID: A viable approach toward optical packet switching," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 7, pp. 1026–1040, Sep. 2003.

[15] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, "Experience in measuring Internet backbone traffic variability:

Models, metrics, measurements and meaning," in *Proc. International Teletraffic Congress (ITC-18)*, Berlin, Germany, Aug.31–Sep.5 2003, pp. 221–230.

[16] K. Fukuda, K. Cho, and H. Esaki, "The impact of residential broadband traffic on Japanese ISP backbones," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 1, pp. 15–22, Jan. 2005.

[17] C. Kattirtzis, E. Varvarigos, K. Vlachos, G. Stathakopoulos, and M. Paraskevas, "Analyzing traffic across the Greek school network," in *Proc. 14th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN'05)*, Crete, Greece, Sep.18–21, 2005.

[18] C. Buyukkoc, P. Varaiya, and J. Walrand, "The $c\mu$-rule revisited," *Advances in Applied Probability*, vol. 17, no. 1, pp. 237–238, Mar. 1985.

[19] G. Koole, "Assigning a single server to inhomogeneous queues with switching costs," *Theoretical Computer Science*, vol. 182, no. 1-2, pp. 203–216, Aug. 1997.

[20] M. I. Reiman and L. M. Wein, "Dynamic scheduling of a two-class queue with setups," *Operations Research*, vol. 46, no. 4, pp. 532–547, Jul.-Aug. 1998.

[21] J.-F. P. Labourdette and A. S. Acampora, "Logically rearrangeable multihop lightwave networks," *IEEE Trans. Commun.*, vol. 39, no. 8, pp. 1223–1230, Aug. 1991.

[22] A. Brzezinski and E. Modiano, "Dynamic reconfiguration and routing algorithms for IP-over-WDM networks with stochastic traffic," *J. Lightw. Technol.*, vol. 23, no. 10, pp. 3188–3205, Oct. 2005.

[23] A. Narula-Tam and E. Modiano, "Dynamic load balancing in WDM packet networks with and without wavelength constraints," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 10, pp. 1972–1979, Oct. 2000.

[24] J. Yates and A. Greenberg, "Reconfiguration in IP over WDM access networks," in *Proc. Optical Fiber Communication Conference (OFC'00)*, vol. 1, Mar.7–10, 2000, pp. 165–167.

[25] I. Baldine and G. N. Rouskas, "Traffic adaptive WDM networks: A study of reconfiguration issues," *J. Lightw. Technol.*, vol. 19, no. 4, pp. 433–455, Apr. 2001.

[26] R. Davey, J. Kani, F. Bourgart, and K. McCammon, "Options for future optical access networks," *IEEE Commun. Mag.*, vol. 44, no. 10, pp. 50–56, Oct. 2006.

[27] L. Xu, H. G. Perros, and G. Rouskas, "Techniques for optical packet switching and optical burst switching," *IEEE Commun. Mag.*, vol. 39, no. 1, pp. 136–142, Jan. 2001.

[28] T. Battestilli and H. Perros, "An introduction to optical burst switching," *IEEE Commun. Mag.*, vol. 41, no. 8, pp. S10–S15, Aug. 2003.

[29] G. I. Papadimitriou, C. Papazoglou, and A. S. Pomportsis, "Optical switching: switch fabrics, techniques, and architectures," *J. Lightw. Technol.*, vol. 21, no. 2, pp. 384–405, Feb. 2003.

[30] L. Choy, "Virtual concatenation tutorial: Enhancing SONET/SDH networks for data transport," *Journal of Optical Networking*, vol. 1, no. 1, pp. 18–29, Jan. 2002.

[31] "Link capacity adjustment scheme (LCAS) for virtual concatenated signals," ITU-T Rec. G.7042/Y.1305, Nov. 2001.

[32] G. Bernstein, D. Caviglia, R. Rabbat, and H. V. Helvoort, "VCAT-LCAS in a clamshell," *IEEE Commun. Mag.*, vol. 44, no. 5, pp. 34–36, May 2006.

[33] P. Bonenfant and A. R. Moral, "Generic Framing Procedure (GFP): The catalyst for efficient data over transport," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 72–79, May 2002.

[34] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun, "IEEE 802.17 resilient packet ring tutorial," *IEEE Commun. Mag.*, vol. 42, no. 3, pp. 112–118, Mar. 2004.

[35] P. Yuan, V. Gambiroza, and E. Knightly, "The IEEE 802.17 media access protocol for high-speed metropolitan-area resilient packet rings," *IEEE Network*, vol. 18, no. 3, pp. 8–15, May-June 2004.

[36] C. Estan, G. Varghese, and M. Fisk, "Bitmap algorithms for counting active flows on high-speed links," *IEEE/ACM Trans. Netw.*, vol. 14, no. 5, pp. 925–937, Oct. 2006.

[37] T. Nguyen, M. Cristea, W. de Brujen, and H. Bos, "Scalable network monitors for high-speed links: a bottom-up approach," in *Proc. IEEE Workshop on IP Operations and Management (IPOM 2004)*, Beijing, China, Oct.11–13,, year =.

[38] H.-A. Kim and D. R. O'Hallaron, "Counting network flows in real time," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM'03)*, San Francisco, USA, Dec.1–5,, year =.

[39] M. S. Borella, J. P. Jue, D. Banerjee, B. Ramamurthy, and B. Mukherjee, "Optical components for WDM lightwave networks," *Proc. IEEE*, vol. 85, no. 8, pp. 1274–1307, Aug. 1997.

[40] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corporation, MA, Tech. Rep. DEC-TR-301, Sep. 1984.

[41] M. Fisher, C. Kubicek, P. McKee, I. Mitrani, J. Palmer, and R. Smith, "Dynamic allocation of servers in a grid hosting environment," in *Proc. 5th IEEE/ACM International Workshop on Grid Computing (GRID'04)*, Pitssburgh, PA, Nov.4, 2004, pp. 421–426.

[42] E. G. Coffman, R. R. Muntz, and H. Trotter, "Waiting time distributions for processor-sharing systems," *Journal of the ACM (JACM)*, vol. 17, no. 1, pp. 123–130, Jan. 1970.

[43] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[44] E. Altman, "Applications of Markov Decision Processes in communication networks: a survey," INRIA, Tech. Rep. RR-3984, Aug. 2000.

[45] D. R. Cox and W. L. Smith, *Queues*. John Wiley, 1961.

[46] J. S. Baras, A. J. Dorsey, and A. M. Makowski, "Two competing queues with linear costs and geometric service requirements: The μc-rule is often optimal," *Advances in Applied Probability*, vol. 17, no. 1, pp. 186–209, Mar. 1985.

[47] J. S. Baras, D. J. Ma, and A. M. Makowski, "K competing queues with geometric service requirements and linear costs: The $c\mu$-rule is always optimal," *Systems and Control Letters*, vol. 6, no. 3, pp. 173–180, Aug. 1985.

[48] H. Levy and M. Sidi, "Polling systems: Applications, modeling, and optimization," *IEEE Trans. Commun.*, vol. 38, no. 10, pp. 1750–1760, Oct. 1990.

[49] H. Takagi, "Queueing analysis of polling models," *ACM Computing Surveys (CSUR)*, vol. 20, no. 1, pp. 5–28, Mar. 1988.

[50] M. Hofri and K. W. Ross, "On the optimal control of two queues with server setup times and its analysis," *SIAM Journal on Computing*, vol. 16, no. 2, pp. 399–420, Apr. 1987.

[51] O. J. Boxma and D. G. Down, "Dynamic server assignment in a two-queue model," *European Journal on Operational Research*, vol. 103, no. 3, pp. 595–609, Dec. 1997.

[52] Z. Liu, P. Nain, and D. Towsley, "On optimal polling policies," *Queueing Systems*, vol. 11, no. 1-2, pp. 59–83, Jul. 1992.

[53] I. Duenyas and M. P. Van Oyen, "Stochastic scheduling of parallel queues with set-up costs," *Queueing Systems*, vol. 19, no. 4, pp. 421–444, Dec. 1995.

[54] ——, "Heuristic scheduling of parallel heterogeneous queues with set-ups," *Management Science*, vol. 42, no. 6, pp. 814–829, Jun. 1996.

[55] S. C. Borst, "Polling systems with multiple coupled servers," *Queueing Systems*, vol. 20, no. 3-4, pp. 369–393, Sep. 1995.

[56] G. D. Down, "On the stability of polling models with multiple servers," *Journal of Applied Probability*, vol. 35, no. 4, pp. 925–935, Dec. 1998.

[57] A. Narula-Tam, S. G.Finn, and M. Medard, "Analysis of reconfiguration in IP over WDM access networks," in *Proc. Optical Fiber Communication Conference (OFC 2001)*, vol. 1, Mar.17–22, 2001, pp. MN4–1 – MN4–3.

[58] A. Bianco, J. M. Finochietto, G. Giarratana, F. Neri, and C. Piglione, "Measurement-based reconfiguration in optical ring metro networks," *J. Lightw. Technol.*, vol. 23, no. 10, pp. 3156–3166, Oct. 2005.

[59] I. Baldine and G. N. Rouskas, "Dynamic load balancing in broadcast WDM networks with tuning latencies," in *Proc. INFOCOM'98*, vol. 1, Mar.29–Apr.2 1998, pp. 78–85.

[60] I. Alfouzan and A. Jayasumana, "Dynamic reconfiguration of wavelength-routed WDM networks," in *Proc. Local Computer Networks (LCN 2001)*, vol. 1, Tampa, FL, USA, Nov.14–16, 2001, pp. 477–485.

[61] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 2005.

[62] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2000.

[63] T. Osogami and M. Harchol-Balter, "A closed form solution for mapping general distributions to minimal PH distributions," in *Proc. International Conference on Performance Tools (TOOLS'03)*, Urbana, IL, Sep. 2003, pp. 200–217.

[64] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2000.

[65] P. G. Harrison and W. J. Knottenbelt, "Passage-time distributions in large Markov chains," in *Proc. ACM SIGMETRICS*, Marina Del Rey, CA, Jun. 2002, pp. 77–85.

[66] J. Abate and W. Whitt, "Numerical inversion of Laplace transforms of probability distributions," *ORSA Journal on Computing*, vol. 7, no. 1, pp. 36–43, 1995.

[67] V. Jacobson, "Congestion avoidance and control," in *Proc. ACM SIGCOMM'88*, Stanford, CA, Aug. 1988, pp. 314–329.

[68] C. V. Hollot, V. Misra, D. Towsley, and W. Gong, "Analysis and design of controllers for AQM routers supporting TCP flows," *IEEE Trans. Autom. Control*, vol. 47, no. 6, pp. 945–959, Jun. 2002.

[69] D. X. Wei, C. Jin, S. H. Low, and S. Hegde, "FAST TCP: Motivation, architecture, algorithms, performance," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1246–1259, Dec. 2006.

[70] L. S. Brakmo and L. L. Peterson, "TCP vegas: End to end congestion avoidance on a global Internet," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 8, pp. 1465–1480, Oct. 1995.

[71] S. Floyd, "HighSpeed TCP for large congestion windows," IETF RFC 3649, Dec. 2003.

[72] T. Kelly, "Scalable TCP: Improving performance in highspeed wide area networks," *ACM SIGCOMM Computer Communication Review*, vol. 33, no. 2, pp. 83–91, Apr. 2003.

[73] The network simulator – ns-2. [Online]. Available: http://www.isi.edu/nsnam/ns

[74] T. Cui and L. Andrew. FAST TCP simulator module for ns-2, version 1.1. [Online]. Available: http://www.cubinlab.ee.mu.oz.au/ns2fasttcp