MIRROR: IP MULTICAST FOR OPTICAL BURST-SWITCHED NETWORKS

Antônio Jorge G. Abelém and Michael A. Stanton

Abstract - New research perspectives opened up by the combination of IP and WDM technologies present an excellent opportunity for reformulating certain aspects of multicast transmission, bringing them more in line with the needs of future generations of IP internetworking. This paper analyses MIRRORS, which proposes modifications to traditional IP Multicast in order to improve its scalability as a function of the number of simultaneously active groups, as well as making it more appropriate for use in optically switched networks. In this analysis, MIRRORS is compared with other major multicast alternatives, using such parameters as: information state requirements, control overhead, cost of packet forwarding and cost of the multicast distribution tree.

Keywords: High-speed networks, multicast in optical networks, IP over WDM, optical burst switching (OBS), Generalised MPLS (GMPLS).

Resumo - Os recentes avanços na tecnologia de transmissão óptica, mais especificamente na multiplexação por comprimento de onda ("Wavelength Division Multiplexing-WDM"), aliados à consolidação do IP como protocolo dominante das redes convergentes, vêm oferecendo novas perspectivas para as futuras gerações de inter-redes. Este trabalho faz uso da evolução dessas tecnologias para analisar um conjunto de adaptações à difusão seletiva, em especial ao IP Multicast, denominado MIRROR ("Multicast IP para Redes baseadas em Rajadas Ópticas Rotuladas"). A proposta MIRROR sugere modificações e adequações para tornar o IP Multicast menos complexo, mais escalável em relação ao número de grupos ativos simultaneamente e mais adequado às redes baseadas em comutação óptica. Na análise, compara-se a MIRROR com as principais alternativas propostas ao IP Multicast, confrontando os parâmetros: requisitos de informações de estado, custo com informações de controle, custo de encaminhamento dos pacotes e custo da árvore de multiponto.

Palavras-chave: redes de alta velocidade, difusão seletiva, IP sobre WDM, comutação de rajadas ópticas (OBS), MPLS Generalizado (GMPLS).

1. INTRODUCTION

The exponential growth of the Internet in recent years, consolidating IP as the leading candidate for the dominant protocol of convergent networking, allied to recent advances in optical technology, most specifically in Wavelength Division Multiplexing (WDM), have made IP over WDM be seen as the basis for a future generation of internetworks [1]. In this evolving situation, multicast transmission has come to be seen as a fundamental aspect of the development of future networks, for reasons which include its being an appropriate paradigm for group-based multimedia applications, and its capacity for more efficient use of network resources [2]. On the other hand, any proposal for using IP multicast over WDM networks must take into consideration not only not only discussions of multicast, but also recent advances in optical networking, and the interaction between IP and the optical transport network.

In order to exploit the theoretical total capacity of optical fibres (~50 Tbps) by using WDM, we have to augment current switching technologies, based on electronic switching with optical interfaces, known as OEO (Optical interface - Electronic switch - Optical interface). OEO equipment limits the capacity of networks, since they are unable to perform switching at much more than a few tens of gigabits per second (10 Gbps for IP routers and 40 Gbps for SONET/SDH switches). In order to switch message flows at higher rates than these, we need to change to optical switching (OOO equipment), potentially much faster, since conversions between optical and electronic domains are eliminated. On the other hand, new difficulties arise in optical switches: both the reading and processing control bits (headers) at very high speed, and the storing of photons, either during header processing or in order to perform statistical multiplexing at output interfaces.

There are three well-known approaches to dealing with these questions, which have been discusses, for example, in [3]. These are lambda switching (optical circuit switching), optical burst switching (OBS) [4], and optical packet switching (OPS) [5]. Of these the OBS paradigm has been receiving much attention from research workers and professionals, as it displays a number of advantages when compared with the alternatives of lambda switching and optical packet switching.

In spite of having been the subject of intensive research in recent years [6], the Internet multicast model, known as IP Multicast [7], is still the subject of much debate and questioning, even though more than ten years have already passed since it was originally proposed [6] [8]. Apart from the widely known and long-studied problems of the lack of effective access control of group membership and of a multicast address allocation scheme, there persists the need for all routers belonging to the multicast tree, whether they be border or internal routers, to maintain state information relating to the multicast distribution tree of each individual group. In the context of IP over WDM, implementing multicast directly

Antônio Abelém is professor of the Computer Science Department of Universidade Federal do Pará (UFPA), Belém, PA, Brazil, and Michael A. Stanton is professor of the Computing Institute of the Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil.
E-mails: abelem@ufpa.br, michael@rnp.br.

at the WDM layer has a number of advantages. In the first place, with knowledge of the physical topology, which may be different from that seen by the upper layers, we can build more efficient multicast distribution trees. Secondly, because of the inherent capacity of light beam splitting of some optical switches, it can be more efficient to replicate optically an entire light beam than to replicate individual IP datagrams electronically. Thirdly, multicast at the WDM layer offers a greater degree of transparency in terms of bit rate and coding format. On the other hand, multicast at the WDM layer has to confront a number of new challenges, such as not all nodes may be capable of splitting light beams or performing wavelength conversion.

As a consequence, it is becoming apparent that new research perspectives are opening up at the intersection of optical networks, specifically WDM networks, with IP internetworks, offering a great opportunity to analyse and reformulate some aspects of IP multicast. Thus the authors presented the MIRRORS proposal [9], with adaptations to IP multicast to make it more scalable with respect to the number of simultaneously active groups, and more appropriate for optically switched (OBS) networks. MIRRORS exploits the advantages of new and promising approaches, such as optical burst switching (OBS) and Generalised MultiProtocol Label Switching (GMPLS) [10]. Basically, the MIRRORS proposal re-examines the requirement that all routers in the multicast distribution tree maintain tree state information, as well as suggesting adaptations in the way multicast paths are established when label switching is used.

This article presents an analysis of the changes proposed in [9]. Comparisons are made with the main alternatives that have been proposed to traditional IP multicast, using such parameters as: state information requirements, the overhead of control information, the cost of packet forwarding and the cost of the multicast tree.

Section 2 presents related work. In Section 3 there is a summary of the MIRRORS proposal, emphasising alterations suggested to the traditional model of IP multicast. Section 4 analyses the MIRRORS proposal, comparing it with major alternatives. In Section 5 we present our conclusions and suggest future work.

2. RELATED WORK

Amongst recent proposals for modifying IP multicast, the ones most discussed are those which have sought to simplify the model and to improve its scalability with reference to the number of simultaneously active groups. The proposals that have stood out are: EXPRESS [11] and its successor SSM ("SSM - Single Source Multicast") [12], REUNITE ("RECursive UNIcast TrEes") [13] and related improvements [14], as well as XCAST ("eXplicit multiCAST") [15].

The EXPRESS proposal consists in the adoption of a channel-based model, in which each multicast group has a single source, and can therefore be identified by (S, G), where 'S' is the source IP address and 'G' the class D IP address of the group. As each group has only one source, EXPRESS has no need to consider the added complexity of shared trees (*, G), as used in traditional IP multicast [7], particularly in relation to inter-domain management. Since it presented simple and efficient solutions to important questions, the EXPRESS proposal has received a lot of attention in the networking community, and a specific IETF working group called SSM (Single Source Multicast) has been set up [12]. Nevertheless, this approach maintains multicast group state information in all routers in the distribution tree, which reduces the scalability of the model as a function of the number of simultaneously active groups.

In an attempt to address this question of scalability, some workers have suggested approaches based on the notion that multicast group state information need only be kept at branching nodes of the distribution tree [13] [14]. Amongst these, REUNITE was well received as its multicast distribution implementation was based on the unicast routing infrastructure. REUNITE also works with a single source per group, although it does not use class D IP addresses. Instead of this, group identification and data forwarding are based on unicast IP addresses. Outing information is separated into two tables, one for multicast control and the other for multicast forwarding. Routers which simply forward packets to a particular group maintain a group entry only in the control table, whereas branching nodes maintain information in their forwarding table. Such information is used at branching nodes to create, recursively, copies of data packets. Such copies have their destination addresses altered, so that all group members receive a copy of the data. However, REUNITE encounters difficulties in building multicast distribution trees when unicast routing is not symmetric [14]. HBH builds on the basic ideas of REUNITE, and suggests modifications to deal with the problems of asymmetric routing.

Another recent proposal for dealing with scalability as a function of the number of simultaneously active groups is XCAST, which also been much commented [15]. XCAST completely does away with the traditional schemes of signalling within a session and of maintaining group state information in the multicast routers. In fact, XCAST uses neither a group management scheme nor a multicast routing protocol. Packet distribution to receivers is performed entirely using unicast transport, and destination IP addresses are kept only by the source. Since each packet carries the addresses of all destinations in the downstream subtree, XCAST tends only to be used for small groups.

3. A SUMMARY OF THE MIRRORS PROPOSAL FOR IP MULTICAST IN OPTICAL BURST SWITCHED NETWORKS

As pointed out in the introduction, the use of IP over WDM presents a good opportunity for analysing and improving a number of features of the traditional model of IP multicast. With this end in mind, the MIRRORS proposal [9] suggests alterations to the way in which IP multicast distribution trees may be built and maintained, to make them more scalable with respect to the number of active groups, and more appropriate for use in optically switched networks.

MIRRORS also deals with a number of questions relating to signalling and control, as well as to protection and restoration. In both these cases, we suggest changes to ex-

isting schemes, in order to bring them more into line with our proposal [9]. We shall, however, restrict ourselves only to describing the proposed alterations to IP multicast, as only these aspects will be evaluated in this article. Specific analysis of modifications of signalling and control, and protection and restoration will be dealt with in future work.

3.1 THE REFERENCE MODEL

The reference model considered here consists of IP/MPLS[1] routers connected by optical internets, capable of multicast transport through dynamically switched light paths (Figure 1). These optical internets are based on the paradigms OBS and MPLS. The choice of OBS technology is due not only to its greater efficiency, as it is unnecessary permanently to dedicate lambdas to each branch of the multicast distribution tree, but also to its greater suitability to IP over WDM networking, since redundancy is minimised, thus increasing the efficiency of these kinds of network.

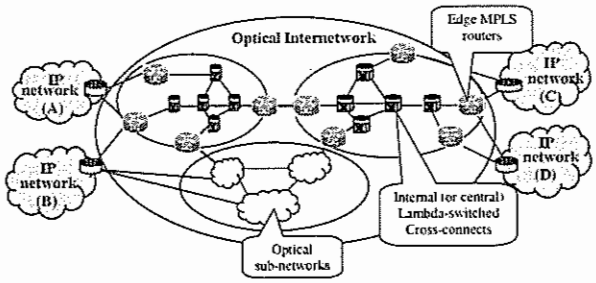

Figure 1. Network model adopted in MIRRORS.

The adoption of MPLS, on the other hand, is a consequence of its simplification of the definition and maintenance of the optical (burst-switched) layer, especially in such matters as those relating to the definition of interfaces, to address allocation and resolution, and to interoperation with the upper layers of traffic management.

We suppose that an optical internet consists of multiple optical networks, possibly under independent administration. Each optical network could be made up out of one or more subnetworks of optical label-switching devices (so-called Lambda-Switched Cross-connects - LSCs), interconnected by optical links in a general mesh topology. Such LSCs are incapable of processing IP packets, and only some of them are even capable of wavelength conversion. For reasons of simplicity, we will suppose there exists one-to-one correspondence between IP controllers and LSCs.

In this labelled OBS (LOBS) architecture, the nodes of the network are classified in two groups: the internal nodes and the edge nodes. The internal nodes perform the switching of bursts based on labels and avoid temporary buffering. The edge nodes possess the electronic functionality appropriate for IP routers and are responsible for the process of burst assembly.

Additionally, we consider an integrated control model, as suggested in [16], where the optical and IP domains are managed in a unified manner, with just one instance of a routing protocol and one control plane, based on MPLS.

Finally, so far as the capacity for multicasting is concerned, the model adopted is of networks with sparse splitting capacity, where only some of the switches are able to perform multicast switching. Another relevant consideration is that, as building multipoint to multipoint trees is complicated in the optical domain, the model proposed here uses only source-based distribution trees, which, as well as being more appropriate to the optical domain, are generally built in a more optimised form.

3.2 MULTICAST DISTRIBUTION TREE AND ROUTING

As was pointed out in the introduction, in addition to existing problems such as its poor scalability, the use of IP multicast in WDM networks presents new challenges for building multicast distribution trees. One of the most important of such challenges is that not all nodes are capable of splitting light beams. Another important consideration is that, in accordance with proposals to offer Differentiated Services (DiffServ) and traffic engineering (MPLS), only border nodes maintain pertinent state information. In consequence, even those internal nodes with the capability of splitting light beams are unable to do this in an intelligent way, as they do not have access to sufficient information. MIRRORS proposes an alternative approach for solving such problems, with the bonus of being more appropriate for optical networks.

Rather than keeping group state information in all nodes of the distribution tree, as in traditional IP multicast, such information will only be kept in the border routers that form part of the distribution tree in each domain. As our reference model adopts the OBS paradigm, information about the distribution tree is encapsulated in the burst control packets (BCPs), which are processed electronically in each network node. In addition, as MIRRORS also is based on use of MPLS, the BCPs are incorporated in new messages of the MPLS control protocols (such as CR-LDP or RSVP), which are exchanged during the process of setting up the label-switched paths (LSPs). From this point on, the optical bursts are label-switched along a pre-established multicast tree, offsetting any additional overhead caused by the encapsulation scheme.

Finally, as a result of the changes we are proposing and of the new challenges in multicast in the context of WDM, it becomes clear that the routing protocols used nowadays in IP Multicast, such as PIM-SM, would not be the most appropriate. For our model, the most appropriate routing algorithms would be those based on knowledge of network topology, capacity, and resource availability. Such information may be stored and used either in a centralised scheme, or in a distributed scheme, using a link state protocol like MOSPF [17]. Our suggestion, which is in accordance with a proposal of the IETF [16], is the adoption of a link state protocol with appropriate alterations for the optical context. It should be mentioned that such proposals already exist [18]. In our extensions, the link state database will identify which internal nodes are capable of splitting light beams, and this database would be stored only in the edge routers of the domain.

[1] Henceforth, we use the term MPLS as a generic way of referring to the use of MPLS or of one of its successors, MPλS or GMPLS.

In the MOSPF routing protocol [17], each router which serves members of a given group should communicate this fact in link state announcements (LSAs) sent to all other routers in the same domain. Such "group-membership-LSA" messages indicate which transit nodes (the router itself, or any directly attached transit networks) must not be pruned during the building of the Shortest Path Tree (SPT). The need to notify all domain routers of such information, whether or not they belong to the distribution tree, is due mainly to the traditional model of IP multicast, in which a group can have several sources. Because of this, if any multicast router in the domain receives packets for a particular group, it has to be capable of building the appropriate SPT.

In MIRRORS, on the other hand, the multicast model is more restrictive, with the adoption of the channel scheme, where each group has only one source (see Section 3.1). As it is already known *a priori* who will be the group's source, it will be unnecessary to send "group-membership-LSAs" to all routers in the domain. In principle, it will only be necessary to transmit such LSAs to border routers offering direct or indirect connectivity to the source or to the source domain, i.e. to those border routers nearest to the source, and which can act as a root of the multicast tree in the domain. If a border router for any reason were to receive traffic for a particular group, and it does not hold any membership state for this group, it should seek information about the group from other border routers in the same domain.

4. AN ANALYSIS OF THE PROPOSAL

In this analysis, we evaluate the efficiency of the adaptations to IP multicast suggested in MIRRORS. First, we present an appreciation of OBS as a switching paradigm. After, we compare MIRRORS with the principal alternative proposals to IP multicast which were identified in Section 2. This comparison has been carried out both by theoretical analysis, and also by Simulations using NS ("Network Simulator") [19], where we have developed a prototype of MIRRORS. In both cases, the parameters which have been selected for comparison include the following: state information requirements, control information overhead, packet forwarding cost, and distribution tree cost.

4.1 EVALUATION OF OBS

The OBS paradigm has been extensively studied and tested in recent years because, as was pointed out in Section 1, it possesses a number of characteristics and functionalities which give it a competitive advantage in several respects, when compared with lambda switching and optical packet switching (see Table 1). Among its main virtues, OBS displays better bandwidth utilisation, lower channel set-up latency and greater flexibility than lambda switching, whilst at the same time requiring a simpler implementation than for optical packet switching [3]. The main problem with OBS is that it is still relatively new and unfamiliar, and it is not completely clear what will be the impact on network performance of its unreliable signalling scheme. Nevertheless, a number of studies have recently been published containing encouraging results, not only for point-to-point communication [20] [21], but also for multicast [22].

Optical Switching Paradigms	1	2	3	4
Lambda	Low	High	Low	Low
Burst (OBS)	High	Low	Medium	High
Packet/Cell	High	Low	High	High

1 - Bandwidth Utilisation
2 - Channel Set-up Latency
3 - Difficulty of Implementation
4 - Adaptability (traffic and faults)

Table 1. Comparison between optical switching paradigms.

In order to obtain numerical results which can better demonstrate the benefits offered by a burst-switched network, we decided to study how bandwidth utilisation is affected by the OBS paradigm. Thus, following [23] and [20], we defined by the following expression the total delay in burst-switched networks, neglecting any contribution from the access networks:

$$D_{total} = D_{edge} + D_{(signal\ +\ propag)} + (L_{burst}/Txbit_{core}), \quad (1)$$

where D_{edge} is the delay suffered at the network edge, and corresponds to the time a burst spends waiting in a buffer for a free channel (wavelength) to be allocated. Thus, queuing delay of arriving packets is bounded by D_{edge}. Additionally, $D_{(signal\ +\ propag)}$ is the signalling delay for channel set-up, including propagation delay, $Txbit_{core}$ represents the bit-rate (channel capacity) in the optical network and L_{burst} the burst-size.

In order to investigate the limiting case, we suppose that the burst-size (L_{burst}) grows linearly with the edge delay (D_{edge}), as is the case with constant bit-rate (CBR) traffic [23]. In other words:

$$L_{burst} = D_{edge} \times Txbit_{edge}, \quad (2)$$

where $Txbit_{edge}$ represents the average bit-rate across the external interfaces of the optical network.

A useful parameter for studying the bandwidth utilisation is the average channel (wavelength) occupancy time [20], called here COT, and defined by:

$$COT = D_{(signal\ +\ propag)} + (L_{burst}/Txbit_{core}). \quad (3)$$

Substituting (2) in (3), we get:

$$COT = D_{(signal\ +\ propag)} + [D_{edge} \cdot (Txbit_{edge}/Txbit_{core})]. \quad (4)$$

Another parameter which clearly shows the benefits of dynamic allocation of wavelengths is the channel utilisation (U) [20], which represents the efficiency with which the channel bandwidth is being used, and is defined as the ratio between

the effectively used channel bandwidth and the transmission capacity of the central nodes of the network:

$$U = (Bw\_\lambda / Txbit_{core}), \quad (5)$$

where

$$Bw\_\lambda = L_{burst}/COT. \quad (6)$$

Substituting (4) in (6) and the resulting expression for Bw_λ in (5), we obtain:

$$U = D_{edge} / \left[(A_{Txbit} \cdot D_{(signal\ +\ propag)}) + D_{edge} \right], \quad (7)$$

where $A_{Txbit} = Txbit_{core}/Txbit_{borda}$. Rather than just the ratio between bit-rates of the core and the external interfaces of the optical network, A_{Txbit} represents the "acceleration" of the bit-rate, when the bits migrate from the electronic switching domain to the optical switching domain.

Assuming the sum of signalling and propagation delays ($D_{(signal+propag)}$) of the order of 5 ms (in the case of a network of around 1000 km in diameter), Figure 2(a) shows the channel occupancy time (COT), whilst Figure 2(b) shows the channel utilisation (U), both as a function of the edge delay (D_{edge}), for different values of A_{Txbit}.

From Figure 2(a) it can be seen that, the greater the value of A_{Txbit}, the shorter will be the channel occupancy time. In the case of purely optical networks, where $Txbit_{core} \gg Txbit_{edge}$, the greater will be the benefits obtained for a paradigm such as OBS, which permits dynamic channel reallocation, when compared with conventional lambda switching. Just by way of an example, even for reasonably high values of edge delay (we assume 50 ms), Figure 2(a) indicates that the channel occupancy time will be around 10 ms when $A_{Txbit} = 10$.

In a similar way, we note from Figure 2(b) that channel utilisation falls significantly with increasing values of A_{Txbit}. This implies that a dynamic channel allocation scheme, such as OBS, allows better use to be made of network resources, since it permits the channel reutilisation, as soon as a burst transmission is completed.

4.2 A COMPARATIVE ANALYSIS OF MIRRORS

We compare MIRRORS with other well-known proposals from the published literature, already described in Section 2, which discuss the same questions as we do. The parameters used in this analysis are: state information requirements [23], control information overhead [13], packet forwarding cost [23], and distribution tree cost [14]. These parameters were identified as the most utilised, and those which best evaluate questions related to scalability, complexity and optimisation in the use of network resources.

State Information Requirements ($Req_{state}(T)$) permit us to assess the scalability of multicast models as a function of the number of simultaneously active groups. It is measured as the ratio of 'Num_Routers_with_State', the number of nodes (routers) in the distribution tree, T, which maintain state information about the groups, to

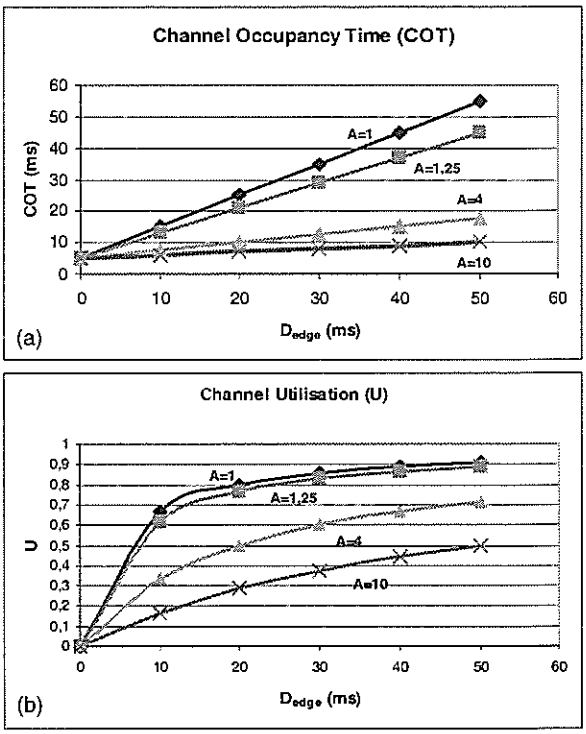

Figure 2. Channel occupancy with the OBS paradigm.

'Total_Num_Routers_in_T', the total number of routers belonging to T:

$$Req_{state}(T) = \frac{Num\_Routers\_with\_State}{Total\_Num\_Routers\_in\_T}. \quad (8)$$

Packet Forwarding Cost measures the work expended by the nodes within the distribution tree to forward the packets to their respective receivers. To do this, one calculates the number of accesses which are made to the routing/forwarding table, and the need for alterations to and/or building of new packet headers by the routers, when packets are forwarded, that is:

$$Cost_{forward} = Access\_Route\_Table + Alt\ \&\ Build\_Pkt\_Head. \quad (9)$$

Control Information Overhead measures the additional bandwidth requirements for control information. These requirements can be represented by different expressions, depending on the scheme being analysed. Because of this we chose initially to represent this cost in generic form:

$$Overhead_{control} = Band\_Reqs\_Control\_Inf. \quad (10)$$

Multicast Distribution Tree Cost is normally measured in two different ways in the literature. Here have adopted the sun of the costs of the individual links which form the distribution trees, to check whether these trees are in fact built using the least cost paths between the source and the destinations.

$$Cost_{mcast\_tree} = Least\_Cost\_Tree \quad (11)$$

4.2.1 STATE INFORMATION REQUIREMENTS

As was pointed out in Section 2, one of the alternatives to the present IP multicast is Single Source Multicast (SSM), which only uses multicast distribution trees rooted in the sender (S, G), avoiding the use of shared distribution trees (*, G) and all their associated complexity. On the other hand, this proposal maintains state information relating to the distribution tree in all routers within the tree. This implies that the state information requirements for this approach always have the value 1, or, in other words:

$$\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) = 1,$$

which implies reduced scalability with respect to the number of simultaneously active groups.

In the case of proposals like REUNITE and HBH, which, apart from in the root and leaf nodes, maintain state information only in branching nodes of the distribution tree, the state information requirements are given by:

$$\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) = (1 + \mathrm{Nodes}_{\mathrm{Branching}} + \mathrm{Nodes}_{\mathrm{Leaf}}) / \mathrm{Total\_Routers\_in\_T},$$

where '1' represents the node (router/switch) which acts as the root of the distribution tree, $\mathrm{Nodes}_{\mathrm{Branching}}$ indicates the number of nodes where there occurs branching of multicast traffic flow, and $\mathrm{Nodes}_{\mathrm{Leaf}}$ indicates the number of nodes at the edge of the multicast tree.

Analysing the above expression, it can be seen that in the vast majority of cases the scheme represented by REUNITE will yield $\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) < 1$. Only in extreme cases, where all internal distribution tree nodes are also branching points, $\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) = 1$. Nevertheless, according to Stoica et al. [13] and Costa et al. [14], it is very rare in practice to encounter multicast trees with the characteristics of these extreme cases.

In the case of the proposal represented by XCAST, where information about the receivers is kept only at the source, the state information requirements are given by:

$$\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) = 1/\mathrm{Total\_Routers\_in\_T}.$$

It should be noted that XCAST has an excellent result for state information requirements, as it does not store any such information in the routers in the distribution tree. In fact, XCAST does not work with this concept. However, as may be seen in the following sections, such a scheme brings in its train grave consequences for the other parameters.

In the case MIRRORS, group information is only stored in those edge routers belonging to the distribution tree. Thus:

$$\mathrm{Req}_{\mathrm{state\_inf}}(\mathrm{T}) = (1 + \mathrm{Nodes}_{\mathrm{Leaf}})/\mathrm{Total\_Routers\_in\_T}.$$

As can be seen from this expression, MIRRORS is fairly scalable in terms of the number of multicast groups, as well as showing good results for state information requirements, without dispensing with multicast management and routing protocols, and this will bring benefits in the analysis of the other parameters.

4.2.2 PACKET FORWARDING COST

In SSM, packet forwarding is identical to traditional IP multicast, where, in general terms, the router consults its multicast forwarding table and, based on the source IP address and the IP address of the multicast group, replicates and retransmits packets through the appropriate interfaces without altering the destination IP address. Thus, the packet forwarding cost will be:

$$\mathrm{Cost}_{\mathrm{packet\_forwarding}} = 1 \text{ access per packet.}$$

Since the approach used by REUNITE and HBH does not use group addresses, the process of packet forwarding by routers at branching nodes is a little more costly. Apart from consulting the forwarding table, packets are not merely replicated, but instead have their headers altered to include the IP addresses of the appropriate destinations. Thus the forwarding cost in this case will be:

$$\mathrm{Cost}_{\mathrm{packet\_forwarding}} = 1 \text{ . access per packet} + \mathrm{N}. \text{ Alter\&Build headers,}$$

where N is the number of outgoing branches of the distribution tree at the branching node.

In the case of XCAST the forwarding cost is still more critical, since here we do not work with group addresses, nor with any kind of group signalling, nor is any kind of multicast state information stored in the routers. Each router at a branching node must make several accesses to its forwarding table in order to retransmit the new packets. In addition, the packets will need to have their XCAST headers rebuilt, with a new (shorter) list of destinations, in the case that they are forwarded by different branches of the distribution tree. Thus:

$$\mathrm{Cost}_{\mathrm{packet\_forwarding}} = \mathrm{ND} \text{ . access per packet} + \mathrm{N}. \text{ Alter\&Build headers,}$$

where ND is the number of destination hosts.

In MIRRORS, forwarding is not critical. The internal tree nodes, in spite of not storing multicast state information, need only access the binary tree (see Section 3) in order to discover through which interfaces the packets should be forwarded. Once this is discovered, the packets will be replicated through the appropriate interfaces, without any need to alter the packet headers. Thus:

$$\mathrm{Cost}_{\mathrm{packet\_forwarding}} = 1 \text{ access per burst.}$$

It may be noted that, even while it reduces the maintenance of group state information at the nodes of the distribution tree, MIRRORS does not raise the packet forwarding cost, unlike XCAST and those schemes that store state information only at branching nodes. This is due to the maintenance within MIRRORS, with some improvements over the original, of such important functions of traditional IP multicast as group addresses and multicast routing.

4.2.3 CONTROL INFORMATION OVERHEAD

Since SSM maintains group information in all routers belonging to the distribution tree, the relevant control informa-

tion overhead is just that relating to "join" and "prune" messages which have to be sent periodically to refresh state information in the routers [12]. The size of each join or prune message is typically around 30 bytes, sent about once a minute [25]. In other words:

$$\begin{aligned}\text{Overhead}_{\text{contr\_inf}} &= \text{messages "join"} + \text{messages "prune"} \\ &= \text{Num\_subnets x 30 bytes.}\end{aligned}$$

In spite of these messages being periodic and not so small, their number does not grow linearly with the number of receivers, as only one join message need be sent per subnet, for any number of receivers.

For REUNITE and HBH, control information overhead is determined basically by the number of signalling messages exchanged between routers. Apart from "join" messages, these schemes both include the "tree" message, and HBH has also the "fusion" message [14]. The "tree" message, sent by the source using IP multicast, contains the information needed to maintain and update the distribution tree structure at the branching nodes of the distribution tree, whilst the "fusion" message are sent by the branching nodes and, together with the "tree" messages, are used in tree construction [14]. Neither REUNITE nor HBH use "prune" messages to indicate leaving a group. As the authors of these two schemes give no details in their papers of the size of signalling messages, especially the two new messages, "tree" and "fusion", it has not been possible to calculate exactly in this case the control information overhead. However, as a result of our analysis, we may deduce that this overhead is greater than for SSM. In other words:

$$\begin{aligned}\text{Overhead}_{\text{contr\_inf}} &= \text{messages "join"} + \text{"tree"} + \text{"fusion"} \\ &> \text{Overhead}_{\text{contr\_inf}}(\text{SSM}).\end{aligned}$$

In XCAST there is no overhead with multicast signalling messages, since, as was mentioned in Section 2, this scheme works neither with group addresses nor with any other multicast functionality. The relevant control information overhead is here contained in packet headers, and consists of the receivers IP addresses, inserted by the source in all packets before transmission. In other words, the control information overhead is:

$$\begin{aligned}\text{Overhead}_{\text{contr\_inf}} &= \text{Num\_Receivers x 32 bits} \\ &= \text{Num\_Receivers x 4 Bytes,}\end{aligned}$$

where Num_Receivers indicates the number of receivers in an XCAST session.

For MIRRORS, the major contribution to control information overhead is due to information contained in the burst control packets, since the internal nodes do not store state information about multicast groups. More specifically, the control information overhead corresponds to the binary tree used to organise and codify the distribution tree. Thus the overhead is

$$\begin{aligned}\text{Overhead}_{\text{contr\_inf}} = \text{K}.\Big(&\lceil \log_2 \text{I} \rceil + \lceil \log_2(\text{K}+1) \rceil + \\ &\lceil \log_2(\text{K}+1) \rceil + \text{G}_\text{m} \Big),\end{aligned}$$

where 'I' represents the number of internal nodes in the domain, of which 'K' belong to the distribution tree and 'G_m' indicates the maximum degree of replication at these nodes.

Analysing the formula for control information overhead in MIRRORS, it can be observed that there is a tendency to grow faster than for XCAST. However, in MIRRORS tree-related information grows as the number of internal nodes in the distribution tree, whereas in XCAST it grows as the number of receivers. That is to say, for MIRRORS, in the worst case scenario, the overhead will be limited by the number of internal nodes in the domain, whilst in XCAST the limiting factor is the maximum number of receivers, a number potentially much greater. Apart from this, as in MIRRORS switching is based on labels, only the control messages sent during the setting up if the LSPs need to contain such additional information. Finally, we should not forget that the control packet will travel in a dedicated (control) channel, and this minimises still further the possible waste of bandwidth.

4.2.4 MULTICAST DISTRIBUTION TREE COST

Firstly, the multicast tree cost will be calculated as a function of the number of copies of the same packet being transmitted across the different links in the network. Analysis of distribution tree construction in the different schemes reveals that SSM, XCAST and MIRRORS will transmit at most one copy of each packet across the links of the tree. The only approach examined which does not always guarantee such behaviour is REUNITE, as here distribution trees are built using the "join" and "tree" messages, which originate in different parts of the tree ("join" messages are sent by receivers, whereas "tree" messages are sent by the source). Since the source addresses data packets to the first member that joined a session, a poor choice of the distribution node may occur in certain cases of asymmetric routing, which produces unnecessary copies of packets on certain links. This incorrect behaviour is corrected in HBH, by adding a further signalling message ("fusion") and requiring the source to address packets to the closest branching node [14].

On the other hand, if we calculate the multicast tree cost as a sum of the costs of the individual links which make up the distribution tree, the results will be completely different. For example, SSM will produce the worst result in most cases, since it builds the distribution tree from the receivers to the source, using Reverse Path Forwarding (RPF) [25], which guarantees the least cost path from receivers to the source, but not necessarily the contrary in the case of asymmetric routing.

The other three schemes analysed here in general produce better results, since, unlike SSM, they build their distribution trees based on the shortest path from source to receivers. The major difference between them is that MIRRORS is more appropriate for working with multicast trees using label-switching. Both XCAST and those schemes which store state information only at branching nodes will encounter difficulties building trees which use label-switching, since their branching nodes need to make alterations in packet headers before forwarding them. To do this requires that internal nodes possess similar functionality to edge nodes, which contradicts a basic tenet of MPLS.

4.2.5 CRITICAL ANALYSIS OF THE RESULTS

The results of the comparative analysis, summarised in Table 1, show that there exists a clear compromise between the maintenance of group state information, packet forwarding costs and control information overhead. It is easily perceived that when there is excessive cuts in the costs of state information and of multicast functionality, as is the case with REUNITE and XCAST, then there is a corresponding increase in the control information overhead and the costs of packet forwarding. On the other hand, if one eliminates just unnecessary state information but retains other multicast features such as group address and group signalling protocols, there is a gain in scalability, without a prohibitive increase in control information overhead or packet forwarding costs.

Such behaviour can be seen in the MIRRORS proposal, which presents rather satisfactory results for three of the four items evaluated. For the only parameter not to present such good results, which was control information overhead, possible negative consequences are minimised since the MIRRORS proposal is based on the LOBS paradigm. This implies that various packets are forwarded with just one header, whilst switching is based on labels, which signifies that only the first burst of a given type of traffic flow need contain an encapsulated form the additional information. Finally, we should not forget that the burst control packet travels in a dedicated control channel, which reduces still further the possible negative consequences of the MIRRORS proposal in terms of wastage of bandwidth.

It can be said that the greatest merit of the MIRRORS proposal has been exactly to look for this balance mentioned in the preceding paragraph, proposing optimisations' which improve the scalability of the model, without losing those multicast functionalities, which with some further improvements, can prevent the control information overhead and the packet forwarding costs from affecting the applicability of the proposal.

In this manner, MIRRORS maintains group state information exclusively in the edge routers, and uses an encapsulation scheme to transport such information to internal nodes. At the same time MIRRORS maintains group addressing and multicast signalling and routing protocols.

As a result, MIRRORS has shown itself to be the most appropriate alternative for the case of future internets based on optical switching, since its characteristics allow the use of label-switching, at the same time as they reduce both the processing performed during forwarding at internal nodes, and the need for temporary packet storage at these nodes, both of which are well-known difficulties for optically switched networks.

4.3 SIMULATION RESULTS AND THEIR ANALYSIS

Simulations were carried out with the aim of confirming the results which were obtained in the comparative analysis. Thus, in the simulations the same parameters were measured as were utilised in the comparative analysis: state information requirements, packet forwarding cost, control information cost and multicast tree cost.

Parameters Proposals	1	2	3	4
SSM	Highest of all	Low	Low	Least in reverse direction
REUNITE & HBH	Medium	High	Medium	Minimum only in HBH
XCAST	Least	Very high	Very high	Least
MIRRORS	Low	Low	High	Least

1 - State Information Requirements
2 - Packet Forwarding Costs
3 - Control Information Overhead
4 - Multicast Distribution Tree Cost

Table 2. A comparison of the alternatives considered.

Since MIRRORS is a proposal most adequate for backbone networks, the simulation topology adopted was chosen to be similar to real backbone networks. Specifically the topology selected was inspired by the ABILENE backbone, as shown in Figure 3.

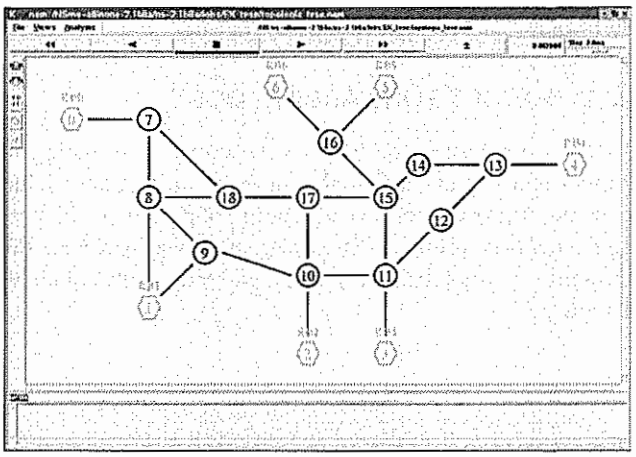

Figure 3. Topology of the network used in the analysis.

The network is made up of 19 nodes, of which 7 act as edge nodes, and the other 12 as internal nodes. The edge nodes (RB0 to 6) are represented as hexagons labelled 'RB'. The internal nodes (7 to 18) are represented by ellipses and numbered conventionally. In order to establish a methodology for group growth, the source[2] was chosen to be RB0, whilst the receivers join the group in the following order: RB1, RB6, RB2, RB5, RB3 and, lastly, RB4. Analyses were made for groups of 2, 3, 4, 5 and 6 receivers.

4.3.1 STATE INFORMATION REQUIREMENTS

The variation, as a function of tree size, of the state infor-

[2]In this article, the terms "group members", "source", "destination" and "receiver" refer primarily to routers and switches.

mation requirements for the four proposals MIRRORS, SSM, REUNITE and XCAST is shown in Fig. 4.

It may be noted that, in this case, the state information requirements for MIRRORS maintained a fairly stable value, around 0.4, for different tree sizes. Although this result is preliminary, it suggests that the simulation results do not contradict the results of the comparative analysis.

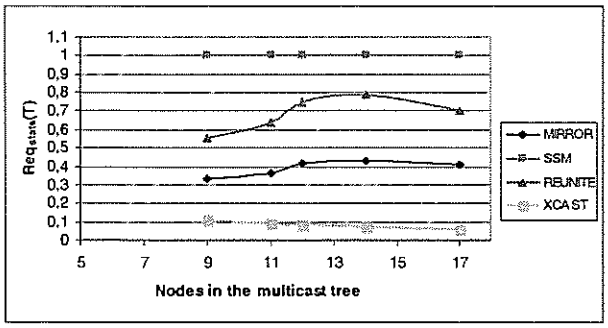

Figure 4. State information requirements for MIRRORS and alternative proposals.

The relative stability of this parameter may also indicate a tendency for an increase in the number of internal nodes in the distribution tree proportional to the number of edge nodes in the same tree. However, such an indication requires further simulations with different topologies in order to be confirmed.

4.3.2 PACKET FORWARDING COST

Here we measured the forwarding cost by estimating the number of accesses made to the route forwarding table added to the number of alterations or building of packet headers during packet forwarding. Such a calculation was performed for different sizes of distribution trees, and the results are shown in Figure 5.

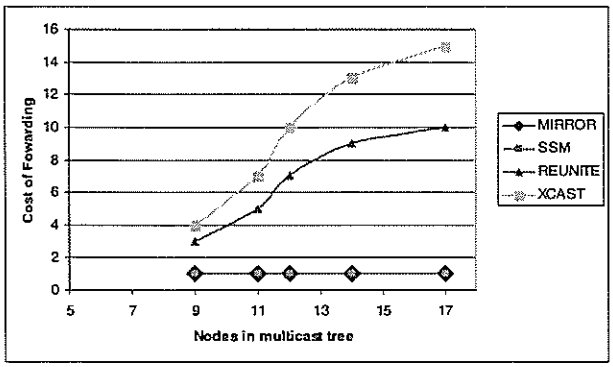

Figure 5. Packet forwarding cost for MIRRORS and alternative proposals.

It should be noted that both REUNITE and XCAST show high forwarding costs, with a tendency to increase as a function of tree size, especially in the case of XCAST. These results suggest similar behaviour to that deduced from the critical analysis for this parameter.

As was mentioned in the critical analysis, REUNITE and XCAST included design choices to implement some basic multicast functions using point to point communication. Although this approach can improve scalability with respect to the number of simultaneously active groups, it can also detract from this same scalability as a result of the increased forwarding cost. On the other hand, proposals like MIRRORS and SSM, which incorporate such basic functions of multicast as group addressing and multicast routing, display low packet forwarding cost, almost independent of tree size.

4.3.3 CONTROL INFORMATION OVERHEAD

As mentioned in the critical analysis above, control information overhead depends on different variables for the different proposals analysed. For instance, in MIRRORS this overhead increases with the number of internal nodes of the distribution tree, whereas in XCAST, it increases with the number of receivers. Because these two proposals are those for which, in principle, the control information overhead is most critical, we decided to confine this part of the study to the two of them. In order to carry this out, in this simulation we measured the overhead due to information contained in packet headers or in control messages for these two proposals, which we consider to be the most relevant in these two cases. The results are shown in Figure 6.

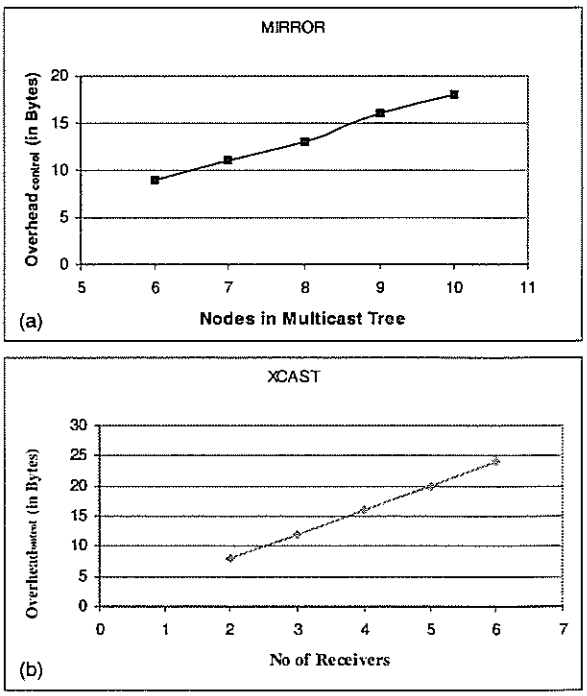

Figure 6. Control information overhead for the proposals MIRRORS and XCAST.

When we observe the evolution of control information overhead of these two proposals for our example network, we can perceive that both of them, in spite of beginning with low values, demonstrate a tendency for a linear or near-linear increase. These preliminary results are also in accordance with the deductions of the critical analysis for this parameter.

However, as mentioned in Section 3, in MIRRORS control information overhead increases with the number of internal nodes in the distribution tree, whereas in XCAST it grows with the number of receivers.

It should be noted that we are not taking into consideration here the fact that in MIRRORS only the first control messages need to contain such additional information, due to the use of labelled optical burst switching. If instead of calculating control information overhead as a function of the number of internal nodes in the distribution tree, or of the number of receivers, we were to calculate this cost as a function of the quantity of control information compared with the quantity of data transmitted, we would expect that MIRRORS would present better results than those of XCAST.

Multicast Distribution Tree Cost

In order to measure the multicast tree cost, simulations were carried out in two circumstances. The first of these used a topology with symmetric links, that is to say, with the cost independent of the direction of the traffic. In the second, we used asymmetric links, which normally are more like those found in real networks [14]. In both cases we computed the multicast distribution tree, using both the lowest cost reverse path, as in SSM, and the lowest cost direct path from source to receivers, as in MIRRORS, REUNITE and XCAST. In the second case, we chose to use only the tree computed in MIRRORS to illustrate the approach based on the lowest cost direct path.

In the case of symmetric routing, the same trees are built by both approaches. This is the expected result as the links have the same costs in both directions. When asymmetric routing was used, the link costs were chosen as in Figure 7. As in the simulator the standard link cost is 1, we have chosen to display only those whose costs were modified, together with the directions of traffic - the remaining costs are all equal to 1.

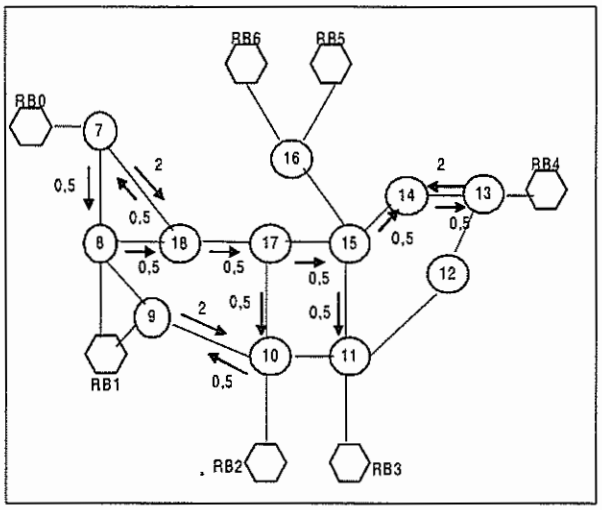

Figure 7. Topology with asymmetric links.

The trees that were built for the case of asymmetric routing in the MIRRORS and SSM proposals are illustrated in Figures 8(a) and 8(b), respectively.

The simulations carried out with asymmetric routing

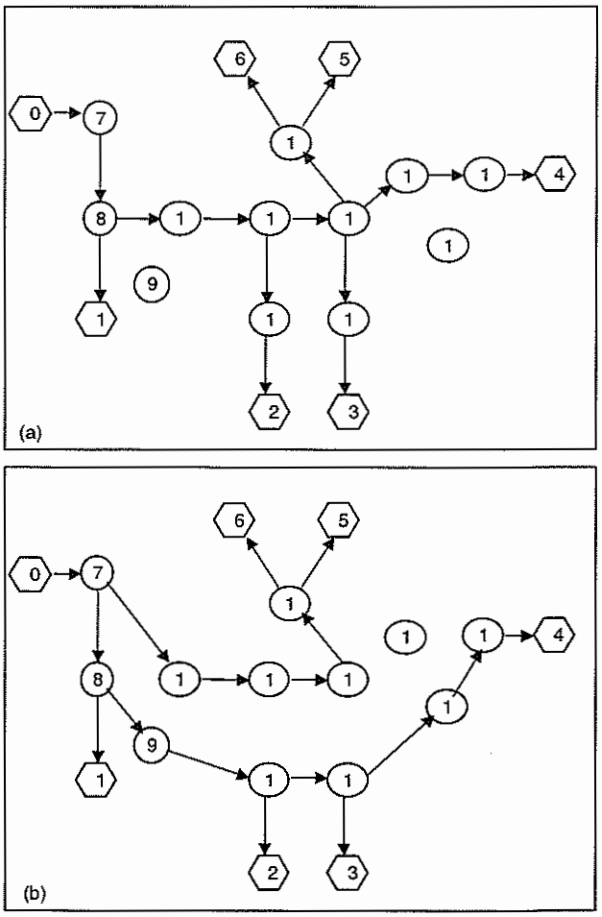

Figure 8. Multicast trees built for the topology with asymmetric links.

clearly confirm the forecasts which were made in the comparative analysis of distribution trees built by the different proposals. Whilst SSM (Figure 7(b)) always establishes the distribution tree using the least cost route from the receivers to the source (the reverse route), the other three proposals build the distribution tree using the least cost route from source to destinations (Figure 7(a)), which is how the traffic will effectively be transmitted.

5. FINAL CONSIDERATIONS AND FUTURE WORK

The MIRRORS proposal shows promising results, with favourable relationship between cost and benefits, shown by the reduction of multicast state information in the distribution tree, and the increased overhead of maintaining and transmitting control information. Additionally, MIRRORS has been shown to be the most appropriate solution, among those examined, for the case of optically switched IP internetworks, based on labelled optical burst switching, since it restricts the need for intelligence and complexity to the network edge, reducing to a minimum the need for processing at internal nodes, and simplifying the setting up of LSPs.

Future work will include conducting tests with the MIRRORS proposal, using traffic engineering and differentiated services techniques.

6. ACKNOWLEDGEMENTS

This work was carried out during Abelém's doctoral studies at the Computer Science Department of the Catholic University of Rio de Janeiro (PUC-Rio), with financial support provided by Capes through the PICDT programme.

REFERENCES

[1] M. Veeraraghavan, M. Karol, R. Karri, T. Moors and R. Grobler. "Architectures and protocols that enable new applications on optical networks." *IEEE Communications Magazine*, vol. 39, no. 3, pp. 118-127, March 2001.

[2] B. Mukherjee. "WDM optical communication networks: progress and challenges." *IEEE Journal on Select Areas in Communications*, Vol. 18, No. 10, pp. 1810-1824, October 2000.

[3] A. Abelém and M. A. Stanton. "Implementing IP over Optical Network" (in Portuguese), Brazilian Symposium on Computer Networks, *Mini-courses*, Chapter 2, pp. 63-123, Búzios, RJ, Brazil. May, 2002.

[4] C. Qiao and M. Yoo. "Optical Burst Switching (OBS) – A New Paradigm for an Optical Internet". *Journal of High Speed Networks (JHSN)*. Vol. 8, No. 1, pp. 69-84, August, 1999.

[5] GUILEMOT, C. et al. "Transparent Optical Packet Switching: The European ACTS KEOPS project Approach". *IEEE Journal Lightwave Tech.* Vol. 16, No. 12, pp. 2117-2134, Dez, 1998.

[6] C. Diot, B. Levine, H. Kassem and D. Balensiefen. "Deployment issues for the IP Multicast service architecture." *IEEE Network*, vol. 14, no. 1, pp. 78-88, January/February 2000.

[7] S. Deering. "Host extensions for IP Multicasting". *RFC 1112*. August 1989.

[8] F.A.R. Barros and M. A. Stanton.. "Are Alterations Needed to the IP Multicast Service Model ?" In: Journal of the Brazilian Computer Society, Vol 7, N° 2, pgs 16- 27, November, 2001. (ISSN 0104-6500)

[9] A. Abelém and M. A. Stanton. "IP Multicast for Optically Switched Networks", *Proceedings of 15° International Conference on Computer Communication (ICCC2002)*, Mumbai, India. August, 2002.

[10] E. Mannie (editor). "Generalized Multi-Protocol Label Switching (GMPLS) Architecture". *Internet Draft*, draft-ietf-ccamp-gmpls-architecture-03.txt. August, 2002.

[11] H. Holbrook et al. "IP Multicast channels: EXPRESS support for large-scale single-source applications." In: *Proceedings of ACM SIGCOMM'99*, September, 1999, pp. 65-78.

[12] H. Holbrook and B. Cain. "Source-Specific Multicast for IP." *Internet Draft*, draft-ietf-ssm-arch-02.txt. March, 2003.

[13] Stoica, et al. "REUNITE: A Recursive Unicast Approach to Multicast." *Proceedings of INFOCOM 2000*. Tel Aviv, Israel. March, 2000.

[14] L. Costa, S. Fdida and O. Duarte. "Hop by Hop Routing Multicast Protocol.". In: *Proceedings of ACM SIGCOMM'2001*, San Diego, EUA. August, 2001.

[15] R. Boivie, et al. "Explicit Multicast (XCAST) Basic Specification." *Internet Draft*, draft-ooms-xcast-basic-spec-03.txt. June, 2002.

[16] B. Rajagopalan et al., "IP over Optical Networks: A Framework." *Internet Draft*, draft-ietf-ipo-framework-03.txt. January, 2003.

[17] J. Moy. "Multicast Extensions to OSPF". IETF *RFC 1584*. 1994.

[18] K. Kompella et al. "Routing Extensions in Support of GMPLS." *Internet Draft*. draft-ietf-ccamp-gmpls-routing-05.txt. August, 2002.

[19] NS. Network Simulator website. URL: http://www.isi.edu/nsnam. Accessed in March, 2003.

[20] M. DUESER, et al. "Bandwidth Utilisation and Wavelength Re-Use in WDM OBS Networks". In: *Proc. IFIP/TC6 5th Conference on Optical Network Design and Modelling*. February, 2001.

[21] BALDINE, I. et al. "JumpStart: A Just-in-Time Signaling Architecture for WDM Burst-Switched Networks". *IEEE Communications Magazine*. February, 2002.

[22] C. QIAO, et al. "WDM Multicasting in IP over WDM Networks". In: *IEEE ICNP'99 Proceedings*. November, 1999, pp. 89-96.

[23] S. Keshav "*An Engineering Approach to Computer Networking*". Addison-Wesley. EUA, 1997.

[24] J. Tian and G Neufeld. "Forwarding State Reduction for Sparse Mode Multicast Communic." *Proceedings of INFOCOM98*. March, 1998.

[25] B. Fenner, et al. Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised). *Internet Draft*, draft-ietf-pim-sm-v2-new-07.txt. March, 2003.

Antônio Jorge G. Abelém graduated in electrical engineering (electronics option) in 1990 from the Federal University of Pará (UFPA), and master's (in 1994) and doctoral (in 2003) degrees from the Catholic University of Rio de Janeiro (PUC-Rio), respectively in Computing Systems and in Computer Networks and Distributed Systems. The title of his doctoral thesis was "Multicast in IP internetworks based on optical networks". He has been a university teacher in the Computer Science Department at UFPA since 1996, and since 2003 he has also been attached to the postgraduate course in Electrical Engineering (PPGEE) at the same university. He has also taught postgraduate extension courses at PUC-Rio, UNAMA (Belém) and UFPA. At UFPA he was also responsible for setting up and coordinating this extension programme in the period 1996-1998. He has acted as a consultant in computer networking for IDESP-PA and for the Brazilian National Research and Education Network, RNP. His areas of interest include computer networking architectures, multicast, optical networks and Quality of Service (QoS).

Michael Anthony Stanton holds BA (1967) and PhD (1971) degrees in mathematics from Trinity College, Cambridge, England. Since 1971 he has held teaching posts at the Aeronautics Technological Institute (ITA) in São José dos Campos, SP (until 1973), at the Catholic University of Rio de Janeiro (PUC-Rio) between 1973 and 1999, and at the Universidade Federal Fluminense (UFF) since 1994. At UFF he holds the post of professor of computer networking in the Computing Institute. He has oriented 29 master's dissertations and 3 doctoral theses, and published almost 100 articles in journals, conference proceedings or as technical notes. Since 2000, he has written over 100 articles published in a regular column, called Virtual Society, which deals with the interactions between information and communications technologies and society, on the website of the media group Estado de São Paulo (www.estadao.com.br/tecnologia). Since 1987, he has been greatly involved in the creation of computer networking infrastructure in Brazil, having belonged to the coordinating body of the Brazilian National Research and Education Network, RNP, between 1990 and 1993, and again since 2001. He was also coordinator of the Rio de Janeiro state academic network between 1989 and 1991. Since 2002, he has been on secondment to RNP as Director of Innovation.