MPLS Working Group
Internet Draft
Expiration Date: February 1999

B. Jamoussi
Nortel Ltd

N. Feldman
IBM Corp

L. Andersson
Bay Networks Inc

August 1998

MPLS Ships in the Night Operation with ATM

<draft-jamoussi-mpls-sin-00.txt>

Status of this Memo

Abstract

Multi-Protocol Label Switching (MPLS) can have several modes of operation over ATM. The MPLS Framework document [1] indicates that MPLS MUST allow 'ships in the night' (SIN) operation with existing L2 switching protocols (e.g., ATM Forum Signaling).

This document identifies the technical requirements that have to be resolved in order to allow for a successful SIN operation between MPLS and ATM Forum protocol stack. Solutions to the various challenges are proposed.

Table of Contents

1. Introduction

In Section 1.1 we summarize the current models of using MPLS over ATM. Section 1.2 highlights the benefits of running MPLS on ATM. In Section 1.3, the requirements for running in Ships in the night (or hybrid) mode are outlined.

1.1 Modes of Operation of MPLS over ATM

MPLS can have several modes of operation over ATM. This section outlines the various operation modes that have been described so far in various MPLS Internet Drafts. We classify MPLS over ATM modes of operation as follows:

1. Use of ATM hardware as a Label-Controlled Interface:

   a. MPLS-Only:

      This mode has been described the most in the current MPLS documentation. A label switching control component is developed to control the ATM switching hardware as described in [2].

   b. Ships in the night:

      This mode of operation has not been described in any level of detail in any of the current MPLS drafts. The only mention was to divide the VPI.VCI space between the ATM control plane and the MPLS control plane [1, 2]. Other SIN operation considerations were out of scope in [2].

2. Use of ATM as a bearer service (B-ISDN):

   a. Use of a VP:

      A VP is established between two LSRs going across an ATM network. The VCI is used as label within the VP [2, 3].

   b. Use of a PVC or SVC and the VCID:

      A PVC or an SVC is established across an ATM network. The VCID is used to identify the two ends of the LSP [3, 4].

The choice of the model of deployment of MPLS on ATM depends on various factors such as the pre-existence of an ATM network.

In the remainder of this document we focus on the ships in the night mode of operation.

1.2 Benefits of Operating MPLS over ATM

Operating MPLS over an ATM network has many benefits described in this section. Many of the ATM hardware features are readily available to MPLS implementations.

- MPLS inherits the ATM hardware label (VPI/VCI) switching. This allows an ATM-LSR to forward packets at very high rates.

- ATM's powerful traffic management features that are already embedded in ATM hardware are available for use by MPLS with the proper hooks in the MPLS control software (e.g., LDP). For example, the queuing and scheduling techniques embedded in ATM hardware allow it to provide many classes of service. This maps nicely to the COS field in the LDP protocol. In addition, to queuing and scheduling, you can also make use of traffic shaping and traffic policing at the boundaries between MPLS domains.

- The use of ATM-LSRs allows for multiple services (voice, video, and data) to be offered on a single platform. Operating MPLS over ATM reduces the number of infrastructures required to provide multiple services. It is hence possible to offer native ATM, a mixture of Frame and ATM services, as well as MPLS using the same infrastructure.

- Protect the capital investment in ATM infrastructure by increasing the utilization of existing ATM network. Allow for an easy migration from IP over ATM to IP over MPLS over ATM. Protect the network management and operational knowledge.

- While the MPLS architectures evolves, ATM offers traffic-engineering solutions to solve today's bandwidth constraints.

1.3 Ships in the Night Mode

Ships in the night (SIN) operation is described in [1]. Section 1.2 of [1] includes the following requirement:

> "MPLS MUST allow "ships in the night" operation with existing layer 2 switching protocols (e.g., ATM Forum Signaling) (i.e., MPLS must be capable of being used in the same network which is also simultaneously operating standard layer 2 protocols)."

SIN operation means that both MPLS and ATM control and routing software are running on the same network. In addition, both MPLS and ATM traffic share the same network infrastructure. However, both protocols are oblivious to each other.

As the deployment of ATM networks continues to grow, it becomes necessary for the successful deployment of MPLS to address the question of how MPLS and ATM would co-exist.

The key advantage of running in SIN mode is that an existing Multi-Service ATM network (running the ATM Forum software stack) can be easily shared with MPLS (running IP routing and LDP). However, the successful integration of MPLS and ATM on the same network requires the following issues to be resolved:

1. VPI.VCI space partitioning
2. Traffic Management
3. Processing Capacity (Memory and CPU)

2. VPI.VCI Space Partitioning

In a label-controlled ATM (LC-ATM) interface, labels are the VPI.VCI fields of the ATM cell. The entire 28-bit VPI.VCI field can be used as a label in a flat hierarchy environment. However, a two level hierarchy can also be achieved by using the VPI and the VCI fields independently; each representing a level of the hierarchy.

ATM Forum stack makes use of the same VPI.VCI field to switch ATM cells within a VP or a VC. Therefore, partitioning of the VPI/VCI space (label space) between ATM and MPLS is necessary.

This issue is easily resolved through configuration. Two VPI.VCI pools (Label-pool) are configured at start-up time. A Label-pool (a sub-set of the VPI.VCI space) is allocated to ATM and another Label-pool to MPLS. Each system only allocates VPI.VCI resources from its

own Label-pool.

In keeping with the idea that both ATM and MPLS control planes are oblivious to each other, it's necessary to avoid having one control plane infer parameters based on the configuration of the other control plane. Therefore, the division of label space between ATM and MPLS needs to be pre-configured per interface. The exchange of the valid VPI.VCI range between two adjacent LSRs/ATM switches is done as follows:

- MPLS uses the LDP to negotiate the range of valid VPI.VCI labels. This information is exchanged during the initialization phase of the LDP session between two peer LSRs as specified in LDP [5].

- ATM uses the ILMI channel to negotiate the valid VPI.VCI range as specified in [6].

Both LDP and ILMI are currently defined to take the intersection of the VPI.VCI range of two adjacent LSRs and ATM switches respectively.

The choice of the boundary between ATM and MPLS is a Network Engineering exercise that takes into account the number of ATM VCs that are expected, whether VC-merge is supported on the interface, the level of stream granularity, among possibly other parameters.

3. Traffic Management

Traffic Management (TM) is one of the key components of MPLS. Currently, TM is proposed to rely on explicit routing (ER), Class of Service (COS) differentiation, and bandwidth reservation. [7] provides an overall framework for TM in MPLS.

In [8], RSVP is used to setup ERs and reserve bandwidth. However, in [5], the LDP is used to setup the ER and to reserve bandwidth for the LSPs.

This Section defines the necessary Traffic Management building block for providing CoS support in MPLS. In addition, it extends the procedure defined in [5] by proposing a simple but effective Label Switched Path Admission Control Mechanism that allows for high statistical multiplexing gain while providing a bandwidth guarantee.

In order for MPLS to provide effective Class of Service (CoS) differentiation and guarantees, a set of traffic management function have to be implemented by the LSRs. These functions includes traffic shaping, traffic policing, queuing and scheduling, and LSP admission control.

Traffic shaping is the TM function invoked at the egress point of a network to limit the rate of traffic down to a specific value. Traffic shapers often have a buffer that absorbs traffic bursts beyond the shaping rate. Traffic shaping ensure that the outgoing traffic rate from LSR1 to LSR2 (shown in Figure 1) does not exceed the bandwidth agreement between ISP1 and ISP2. This function is usually needed at the boundary between two administrative domains (e.g., between ISP1 and ISP2).

```
            |                                 |
  LSR1      |      LSR2             LSR3      |         LSR4
 +------+   |     +------+         +------+   |      +------+
 |      |   |     |      |         |      |   |      |      |
 |      |   |     |      |         |      |   |      |      |
 |    S |---+-----| P  A |---------|    S |---+------| P    |
 |    A |   |     |    Q |         |    A |   |      |      |
 |    Q |   |     |      |         |    Q |   |      |      |
 |      |   |     |      |         |      |   |      |      |
 +------+   |     +------+         +------+   |      +------+
            |                                 |
     ISP 1  | ISP 2                           | ISP 3
            |                                 |
            |                                 |
         Boundary                          Boundary
```

Legend

S Shaping
P Policing
A Admission Control
Q Queuing and Scheduling

Figure 1. MPLS Traffic Management Strategy

Traffic policing is the TM function invoked at the ingress point of a network to ensures that only the agreed upon traffic rate is allowed to consume network resources. Traffic policing is invoked by a network service provider to protect its network resources (e.g., bandwidth) from traffic that exceeds the traffic contract. Traffic policing ensure that the incoming traffic rate from LSR1 to LSR2 does not exceed the bandwidth agreement between ISP1 and ISP2. This function is usually needed at the boundary between two administrative domains (e.g., between ISP1 and ISP2).

A queuing and scheduling system is required in order to provide differentiated service levels. For instance three emission and four discard priorities provide twelve different service levels. The

emission priorities affect the delay observed by the traffic. The discard priority affects the loss observed by the traffic. Delay sensitive traffic uses the higher emission priority queue than less delay sensitive traffic. Loss sensitive traffic uses the higher discard priority (last one to discard) than the less loss sensitive traffic. A sample queuing and scheduling system is described in Appendix B.

Traffic shaping, policing, and queuing and scheduling are functions often developed in hardware so that they don't impact data traffic throughput. Most well-designed ATM hardware includes all of these function. Therefore, an MPLS implementation on Label-Controlled ATM interface (LC-ATM), inherits these functions.

Admission control ensures that bandwidth guarantees are met. This function is often implemented in software and invoked at LSP setup time. This Internet Draft proposes a simple yet effective mechanism to reserve bandwidth for Explicitly Routed Label Switched Paths. We focus in Sections 3.1-3.2 on LSP admission control which is the procedure used to decide if a request for an LSP can be accepted, based on the network capacity and the attributes of both the requested and the existing LSPs. LSRs must perform LSPAC during the LSP setup process as part of providing bandwidth guarantee.

3.1 Bandwidth Management

An important aspect of SIN operation is bandwidth management. ATM and MPLS traffic share the same network facilities. The support of SIN requires that the interaction between ATM traffic and MPLS traffic is handled carefully.

A multi-service ATM network provides strict bandwidth guarantees to its connections. MPLS Label Switched Paths (LSPs) can also reserve bandwidth through the Label Distribution Protocol as defined in [5].

The bandwidth reservation issue can be resolved in various ways through configuration at start-up time. The following configuration options are possible:

- Hard partitioning between ATM and MPLS
  The aggregate interface bandwidth is partitioned between ATM and MPLS. Two bandwidth pools are created; one for ATM and another for MPLS. Each pool is allocated a percentage of the interface capacity. ATM makes its bandwidth reservations from the ATM pool and MPLS makes its bandwidth reservations from the MPLS pool.

  The choice of the bandwidth pool boundary between ATM and MPLS

is a Network Engineering exercise that takes into account the expected bandwidth requirement of ATM and MPLS. An over-booking factor can also be introduced where the sum of the percentages allocated to ATM and MPLS add to more than 100%.

- Full sharing between ATM and MPLS
  A single bandwidth pool is kept for both MPLS and ATM per interface. As ATM connections or MPLS LSPs are admitted the available bandwidth in the pool gets decremented.

The concept of bandwidth pools can be extended to provide a bandwidth pool per ATM service category and per MPLS class of service in both configuration modes.

3.2 Bandwidth Reservation in LDP

In [5], the following Reservation (RES) object is defined in Section 4.4.4.12:

```
"
 +----------------+-------+-------------------------+----------+
 | OBJECT         | Type  |  Subtype(s)             | Length   |
 +----------------+-------+-------------------------+----------+
 | RES            | 0x0C  |  0x01  Raw Bandwidth    | Variable |
 +----------------+-------+-------------------------+----------+

 SubType = 0x01  Raw Bandwidth

 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                      BW requirement                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BW Requirement

Unsigned 32 bit integer representing the bandwidth, in units of 64,000 bps, that must be reserved for the LSP at the LSR identified in the ERNH Object that contains this Reservation Object."

There is no clear indication on what the value of the BW Requirement field in the RES object should be.

Using a similar terminology to the one used in [9], we define a connection's traffic characteristics as follows:

- R: peak rate,

- b: burst size,
- r: average rate

The BW requirement can be set to the peak rate R, average rate r, or to an equivalent rate (e) of the LSP. Setting the BW requirement to R of the LSP that has an On/Off traffic profile results in more bandwidth reserved than actually needed. However, setting the BW Requirement field to the r would result in high packet loss.

Therefore, in order to determine the BW Requirement field, it is necessary to compute an equivalent rate of the Label Switched Path (LSP). The computation of this rate needs to ensure two objectives simultaneously:

1) Bandwidth guarantees; and
2) Efficient network resource utilization.

There are two ways of making the bandwidth reservation through LDP. In the first option, the equivalent bandwidth is computed only once at the LSR originating the ERLSP based on the traffic characteristics (R, r, and b) of the LSP. The value e is then carried by LDP in the RES object to inform the LSRs along the LSP of the bandwidth requirement. The second option is to extend LDP to carry all three traffic parameters and expect each LSR along the path to make its own computation of the required bandwidth for the given LSP.

An extension to the current LDP specification [5] would be required to support option 2 and to allow for the signaling of more bandwidth parameters (R, r, and b) to more accurately characterize the bandwidth requirements of an LSP.

Appendix A describes an example of an algorithm for computing e.

3.3 Queuing and Scheduling

ATM provides strict QoS guarantees. Hence, MPLS traffic should not interfere with ATM traffic in any way that affect its QoS guarantees. In addition, when MPLS provides bandwidth reservation and guarantees, it is necessary to ensure that native ATM traffic does not interfere with MPLS.

Bandwidth management presented in Section 3 works at the reservation level. However, nodal-mechanisms of each LSR or ATM switch have to be setup such that the QoS guarantees of ATM and MPLS are both met.

The nodal-mechanisms need to ensure that each service stays within its bandwidth allocation limits. Traffic shaping and policing are necessary to limit the throughput of each of the ATM and MPLS

services to the contracted values.

In addition, loss and delay guarantees for both ATM and MPLS have to be met. This requires a careful mapping of ATM service categories and of MPLS Classes of Service on the shared queuing and scheduling system.

Appendix B describes a Multi-Priority System (MPS) presented as an example of a queuing and scheduling mechanism and the mapping of MPLS COSs and ATM service categories to the MPS.

3.4 Alignment with the DS-byte

The Diff-Serv model indicates that traffic conditioning is done at the edge of the network. A codepoint is assigned to define the behavior aggregate of the packet. Within the core, packets are forwarded based on the PHB associated with the DS codepoint. However, when going over an MPLS network where the underlying infrastructure is ATM, the DS codepoint is not accessible to the hardware making the forward decision. Therefore, a mapping between the DS codepoint and an MPLS LSP that would provide the same PHB is required as packets enter the MPLS domain. This means that multiple LSPs are established for a given FEC to accommodate the various DS codepoints.

Currently, only two PHBs are proposed; the DE and the EF PHBs. The [9] draft indicates that the mapping from the DS-byte codepoint to the behavior MUST be configurable. Since in an MPLS domain running on ATM, the mapping from the DS-codepoint to the LSP is only feasible on the edge of the MPLS domain, the mapping has to also be configurable.

The DS-byte includes an in/out bit that indicates whether the packet is within its contracted rate or not. Setting this bit to 'out' means that under congestion, this packet should be discarded before packet that have the bit set to 'in'. The in/out bit in the DS-byte is similar to the CLP bit in ATM cells. Therefore, as part of the mapping function from the DS-byte codepoint to an LSP with a given priority, the In/Out bit should be mapped to the CLP bit of all the cells resulting from the segmentation of the frame.

When going from an MPLS-ATM domain to a Diff-Serv domain, a mapping is also required from the various LSPs to a set of codepoints. In addition, if the CLP bit of any cells of a frame are set, then the "Out" bit needs to be set accordingly.

4. Processing Capacity

When running in SIN mode, two routing and control stacks will reside on the same nodes. ATM requires its signaling, and routing (IISP or

PNNI) software and topology database. MPLS requires its signaling (LDP) and IP routing (RIP, OSPF, or BGP) software and topology database.

Therefore, SIN operation requires that nodal processing capacity (CPU, and memory) is adequate to simultaneously handle ATM and MPLS.

5. Summary

Ships in the night operation support is a required element in MPLS [1]. This document highlights the requirements for SIN operation in MPLS. Solutions are proposed to carefully share various resources including VPI.VCI space, bandwidth, queuing and scheduling, and processing capacity.

The traffic management building blocks that are necessary in order for MPLS to provide CoS differentiation and guarantees are identified. It proposed a simple yet effective mechanism to compute the Equivalent Rate (e) of an Explicitly Routed Label Switched Path to be signaled in the LDP RES object. The e computation allows for high network bandwidth utilization while minimizing packet loss. An admission control mechanism is described based on the e computation.

6. Security Considerations

Security issues are not discussed in this draft.

7. Acknowledgement

The authors would like to acknowledge the valuable comments of Pasi Vaananen, Osama Aboul-Magd, Ken Hayward, and Don Fedyk. In addition, recent MPLS Working Group discussions helped shape some sections of this draft.

8. References

[1] R. Callon, et. al., "A Framework for Multiprotocol Label Switching", draft-ietf-mpls-framework-02.txt, November 21, 1997.

[2] B. Davie, et. al., "Use of Label Switching With ATM", draft-davie-mpls-atm-01.txt, July 1998.

[3] Ken-ichi Nagami, et. al., "VCID Notification over ATM link", <draft-ietf-mpls-vcid-atm-00.txt>, March 1998.

[4] M. Suzuki, "The Assignment of the Information Field and Protocol Identifier in the Q.2941 Generic Identifier and Q.2957 User-to-user Signaling for the Internet Protocol", <draft-ietf-mpls-git-uus-

00.txt>, June 1998.

[5] Anderson, et. al., "Label Distribution Protocol", draft-mpls-ldp-00.txt, March 1998.

[6] "Integrated Local Management Interface (ILMI) Specification Version 4", The ATM Forum technical committee, af-ilmi-0065.000, July 1996.

[7] P. Vaananen, et. al., "Framework for Traffic Management in MPLS Networks", <draft-vaananen-mpls-tm-framework-00.txt>, March 1998.

[8] B. Davie, et. al., "Use of Label Switching With RSVP", <draft-ietf-mpls-rsvp-00.txt>, March 1998.

[9] Y. Bernet, et. al., "A Framework for Differentiated Services", <draft-ietf-diffserv-framework-00.txt>, May 1998.

9. Authors' Addresses

Bilel Jamoussi
Nortel (Northern Telecom), Ltd.
PO Box 3511 Station C
Ottawa ON K1Y 4H7
Canada

EMail: jamoussi@nortel.ca

Nancy Feldman
IBM Corp.
17 Skyline Drive
Hawthorne NY 10532
Phone: 914-784-3254

EMail: nkf@us.ibm.com

Loa Andersson
Bay Networks Inc.
3 Federal Street
Billerica, MA 01821

EMail: andersson@baynetworks.com

Appendix A Example Equivalent Rate Computation

The LSP admission problem is to find a nominal rate, called the equivalent rate (e), for each connection so that the system meets the specified performance objectives. This condition would be achieved as long as the sum of the e values of the accepted LSPs does not exceed the capacity of the designated link.

If the number of admitted LSPs exceeds the number indicated by the bound, then the packet loss will likely exceed the required target. However, if the number of admitted calls is less than this bound, then link bandwidth is wasted.

The estimate of e is computed by taking the average of the minimum and maximum number of LSPs that can be supported by the link.

The minimum number of LSPs, min, that can be admitted without introducing any nodal loss or delay is obtained such that the sum of the peak rates (R) is less than the link rate (L) (min * R < L). Hence min=L/R.

The maximum number of connections, max, is obtained when the sum of the average rates (r=A*R), where A is the source activity (A=r/R), exceeds the link rate (R * max < L), and there is sustained congestion. Hence, max = L / (A*R). When dividing the link rate by the average of min and max, we obtain the equivalent rate (e), given by (EQ1).

$$e = R * [(2*A) / (1+A)] \qquad \text{(EQ 1)}$$

Note that the link rate L simplifies in the EQ 1 and is no longer needed in the computation of e.

After computing (EQ 1), only once per LSP, the resulting e is signaled in the LDP message establishing the LSP. e is compared to the Available Rate (AvR) value for each hop (link) as follows:

```
if (e < AvR)
   accept the LSP;
else
   reject the LSP;
```

Appendix B Example of ATM Service Categories and MPLS COS Mappings

In this appendix, an example of how ATM and MPLS can be mapped in a multi-priority system is presented.

Most well designed ATM hardware provides a way of differentiating the quality of service received by the various ATM service categories (CBR, VBR, UBR, etc.) through various queuing and scheduling techniques.

In order to offer various delay and loss guarantees, ATM hardware often uses multiple emission and discard priorities. Let's say we have 3 emission and 4 discard priorities. This MPS provides 12 different qualities of service as shown in Figure 2. Traffic is emitted in the order from E1 to E3. Therefore E1 has the lowest relative delay and E3 the highest. Under congestion, traffic is discard in the order from D4 to D1. Therefore, D1 has the lowest relative packet loss and D4 the highest.

```
                                                Emission
                                                Priority

 +---------------------------------------------+  ^
 |          |          | CBR      | Control  |  E1
 |          |          |          |          |
 +---------------------------------------------+
 +---------------------------------------------+
 |          | CoS 2    | rt-VBR   | Control  |  E2
 |          |          |          |          |
 +---------------------------------------------+
 +---------------------------------------------+
 | UBR      | nrt-VBR  |          | Control  |  E3
 | CoS 0    | CoS 1    |          |          |
 +---------------------------------------------+
      D4         D3         D2        D1

              Discard Priority
```

Figure 2 Multi-Priority System (MPS)

Let's consider the open-loop ATM service categories (CBR, rt-VBR, nrt-VBR, and UBR). For MPLS, we define three different Classes of Service.

```
CoS 0  Best Effort
CoS 1  Low Loss Guarantee
CoS 2  Low Delay Guarantee
```

Figure 2 also shows the mapping from the ATM service categories to the MPS as well as the mapping of MPLS CoSs to MPS. CoS 0 and UBR have a similar definition ("best effort"). Hence, they share the same emission and discard priority (the lowest). CoS 1 and nrt-VBR are both targeted for low loss guarantee but do not care too much about delay. Therefore, they are mapped to the E3/D3 slot. CoS 2 and rt-VBR require Low delay guarantee. Hence, they are mapped to a higher emission priority than CoS 0/ CoS 1. In order to provide rt-VBR with slightly better loss guarantee than CoS 2, rt-VBR is mapped to a higher discard priority (D2). Finally, CBR requires the best low delay and low loss guarantees. Hence it is mapped to the highest emission and discard priority.

Network control traffic is in the case of ATM the signaling, ILMI, and RCC channels and in the case of MPLS the LDP channel etc.