Policy-Based Load-Balancing in Traffic-Engineered MPLS Networks

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026 [1].

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

1. Abstract

This document considers the application of policy based traffic engineering techniques to load-balancing in MPLS networks. The objective is not to mandate a specific policy, but to (a) provide an example of a policy based approach to a traffic engineering problem and (b) elucidate a range of potential policies related to load balancing as a traffic engineering objective as guidance for the development of a Policy Information Base (PIB) for MPLS networks.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [2].

3. Introduction

Traffic Engineering within IP networks is a broad subject outlined
within the traffic engineering framework document [3]. This draft takes a considerably narrower approach in considering just one aspect of traffic engineering in order to scope the work.

RFC 2702 [4] section 3.2 identifies three fundamental problems in traffic engineering over MPLS networks - mapping traffic to Forwarding Equivalence Classes (FECs), mapping FEC’s to label Switched Paths (LSPs), and mapping LSPs to physical topology. This draft seeks to address aspects of the first two objectives only.

Specifically, this draft focuses on:

a) MPLS as an interesting subset of IP protocols
b) load balancing as a traffic engineering objective
c) Policy based approaches for describing the objectives and constraints of the traffic engineering optimization.

The MPLS architecture draft identifies load balancing as a potential application of MPLS. The load balancing problem in MPLS networks concerns the allocation of traffic between two or more LSPs which all have the same origin and destination.

The notion of policy-based MPLS networks is introduced in [5] and expanded in [6], in conformance with the generic policy work from the policy working group (see e.g. [7]) and the COPS material from the rap working group (see e.g. [8]).

3.1 Terminology

Load Balancing vs Link Bundling

In some cases a pair of LSRs may be connected by several (parallel) links. From the MPLS Traffic Engineering point of view, for the purpose of scalability, it may be desirable to treat all these links as a single IP link – an operation known as Link Bundling (ref. [9]). With load balancing, the load to be balanced is spread across multiple LSPs that in general does not require physical topology adjacency for the LSRs. The techniques are complementary. Link bundling provides a local optimization that is particularly suited for aggregating low speed links. Load Balancing is targeted at larger scale network optimizations.

While load balancing is perhaps commonly thought of as applying between edge LSRs, it could be applied at ANY LSR that provides the requisite multiple LSP tunnels with common endpoints. The Policy Enforcement Point is the LSR at the source end of the set of LSPs with common endpoints. The arriving traffic to be load balanced may be from non-MPLS interfaces or MPLS interfaces. In general, the source end of an LSP may act as a merge point for multiple input streams of traffic.

3.2. Assumptions / Scope

This document considers only policies related to load balancing objectives. While policy based approaches may be applicable to other traffic engineering objectives, the focus here is on load balancing.
The set of LSPs over which the load is to be balanced is pre-defined and the relevant load balancing policies are then applied to these LSPs. Extension of this work to support the creation or deletion of LSPs in response to policies with load balancing objectives is for further study.

This version of the document is primarily focused on MPLS networks without quantitative guarantees on LSP performance in order to reduce the effort involved. Future versions of the document may include increased support for the various QoS notions supported in MPLS networks. By removing the QoS elements from the discussion at this stage, we can concentrate on getting the fundamental operation correct, without debating the details of QoS or admission control algorithms.

By not considering the QoS notions of MPLS, we simplify the policy and functions associated with load balancing. Since only best effort LSPs are considered, we simplify the admission control considerations in the load balancing process. If the LSPs were established with QoS constraints, it would be necessary to determine if the traffic flow sent over the LSP as a result of load balancing fit the profile of the constraints. This adds complexity to the load balancing policy as well as the processing of the LSR performing the load balancing.

While load balancing on a best effort network can be viewed as a simple case, the basic methodologies have a wider applicability when applied to QoS-based LSP selection. Indeed the load balancing case for best effort only traffic has similar problems to that of load balancing a particular traffic class such as that with a particular DiffServ PHB. Bandwidth sharing among classes of service raises some more complex issues that also apply to the placement of traffic into ER-LSPs. As the available capacity for a particular traffic class to a particular destination exceeds the capacity of the (ER-)LSP for that traffic, some action must be taken to get more bandwidth or control access to the LSP. The PEP will have to perform some per traffic class action with the likely result that the best effort traffic on the network will become squeezed in favor of higher priority traffic. The lending of bandwidth between LSPs is a policy issue that must be addressed. The location of the network congestion will have a bearing on a solution. A policy server can initiate a new LSP and map certain flows to this to avoid the congestion point, thereby improving the performance of those flows and reducing the congestion problem. This would, however, require a congestion detection methodology and inter-PDP communication.

4. Policy-based Traffic Engineering

Wright Informational - Expires December 2000 3

Policy-Based Load-Balancing in TE MPLS Networks June 2000

A policy provides a rule of the form:

IF <condition> THEN <action>

Policy based networking is one of a number of mechanisms that could
be used in achieving traffic engineering objectives. While traffic engineering may be considered an optimization problem, policy approaches provide considerable flexibility in the specification of the network optimization objectives and constraints.

4.1 Policy-based Traffic Engineering in the Context of the Traffic Engineering Framework

Within the TE framework's taxonomy of traffic engineering systems, policies may be:
- dependent on time or network state (either local or global state)
- based on algorithms executed offline or online
- stored centrally (e.g. in a directory) or distributed to an engineerable number of policy decision points
- prescriptive or descriptive
- designed for open loop or closed loop network control

We assume network feedback to be an important part of policy-based networking. While network configuration (provisioning) can be performed in an open-loop manner, in general, policy-based networking should imply a closed-loop mechanism. The distribution and performance of the policy system itself is beyond the scope of this draft – we assume that adequate resources can be provisioned to meet the required policy update frequency etc.

The traffic engineering framework identifies process model components for:
- measurement
- Modeling, analysis and simulation
- optimization.

Policies may be used to identify relevant measurements available through the network and trigger appropriate actions. The available traffic metrics for determining the policy trigger conditions are constrained e.g. by generic IP traffic metrics (see e.g. RFC 2679 [10] or RFC 2722 [11]).

Policies provide an abstraction of network resources – a model that can be designed to achieve traffic engineering objectives. Policies can provide a degree of analysis by identifying network problem through correlation of various measurements of network state. A policy based network is not, however, a network simulation.

A set of policies can be designed to achieve an optimization of network performance through appropriate network provisioning actions.

4.2 Policy Based Load Balancing

Load Balancing, as a traffic engineering objective, is not described in great detail within the TE framework draft. There is some mention of it as an example of a relatively "simple" and "fast" computation that may be suitable for online traffic engineering in section 5.2. It is also identified as a traffic engineering objective within
section 6.6 on content distribution (webserver) requirements. In general though, load balancing would seem to fit within section 6.3 of the TE framework draft as an example of traffic mapping.

The relative simplicity of load balancing algorithms is what makes them attractive for the purposes of this draft in illustrating policy approaches to traffic engineering in the context of MPLS networks.

While load balancing optimizations have been proposed for various routing protocols, such approaches complicate existing routing protocols and tend to optimize towards a fairly limited set of load balancing objectives. Extending these towards more flexible / dynamic load balancing objectives seems overly complicated. Hence this draft takes the approach of building on the policy-based networking architecture which provides mechanisms specifically designed to support flexible and dynamic administration.

5. Policy-Based MPLS Network Context for Load Balancing

Figure 1 illustrates a generic policy based network architecture in the context of an MPLS network. In this example, we have two LSPs established:
- LSP #1 that follows the path(abd)
- LSP #2 that follows the path(acd).

In this draft we do not consider the problems of establishing the LSPs. We assume that a variety of mechanisms may be used for either manual (e.g. LSPs provision explicit routes) or automated (e.g. LSPs based on topology driven or data driven shortest path routes) establishment of the LSPs. Future versions of this or another draft may address the issue of establishing LSPs via policy mechanisms (e.g. using COPS push).

The load balancing operation is performed at the LSR containing the ingress of the LSPs to be load balanced. This LSR ((a) in Figure 1) is acting as the Policy Enforcement Point for load-balancing policies related to these LSPs. In this context the load-balancing problem concerns the selection of suitable policies to control the admission of flows to both LSPs.

The admission decision for an LSP is reflected in the placement of that LSP as the Next Hop Forwarding Label Entry (NHFLE) within the appropriate routing tables within the LSR. Normally, there is only one NHLFE corresponding to each FEC, however there are some circumstances where multiple NHLFEs may exist for an FEC.

The conditions for the policies applying to the set of LSPs to be load balanced should be consistent. For example, if the condition used to allocate flows between LSPs is the source address range, then the set of policies applied to the set of LSPs should account for the disposition of the entire source address range.
Figure 1 LSR as PEP

For policy-based MPLS networks, traffic engineering policies would also be able to utilize for both conditions and actions the parameters available in the standard MPLS MIBs i.e.

- MPLS Traffic Engineering MIB [12],
- MPLS LSR MIB [13]
- MPLS Packet Classifier MIB [14]

MIB elements for additional traffic metrics are for further study beyond the scope of this internet draft.

5.1 Load Balancing at Edge of MPLS Domain

Flows admitted to an LSP at the edge of an MPLS domain are described by the set of Forwarding Equivalence Classes (FECs) that are mapped to the LSPs in the FEC to NHLFE (FTN) table.

The load-balancing operation may be considered as redefining the FECs to send traffic along the appropriate path. Rather than sending all the traffic along a single LSP, the load balancing policy operation results in the creation of new FECs which effectively
partition the traffic flow among the LSPs in order to achieve some load balance objective.

Consider as an example, two simple point-to-point LSPs, (a) and (b), with the same source and destination, over which we are to load balance some aggregate FEC (z). The aggregate FEC (z) is the union of FEC (a) and FEC (b). The load balancing policy may adjust the FEC (a) and FEC (b) definitions such that the aggregate FEC (z) is preserved.

5.2 Load Balancing at interior of MPLS Domain

Flows admitted to an LSP at the interior of an MPLS domain are described by the set of labels that are mapped to the LSPs in the Incoming label Map (ILM).

A Point-to-Point LSP that simply transits an LSR at the interior of an MPLS domain does not have an LSP ingress at this transit LSR.

Merge points of a Multipoint-to-Point LSP may be considered as ingress points for the next link of the LSP.

A label stacking operation may be considered as an ingress point to a new LSP.

The above conditions, which put multiple LSPs onto different LSPs, may require balancing at the interior node. The FEC of an incoming flow may be inferred from its label. Hence load-balancing policies may operate based on incoming labels to segregate traffic rather than requiring the ability to walk up the incoming label stack to the packet header in order to reclassify the packet. The result is a coarse load balancing of LSPs (not flows) onto one of a number of LSPs from the LSR to the egress LSR.

5.3 Load Balancing with multiple NHLFEs

The MPLS Architecture identifies that the NHLFE may have multiple entries for one FEC. Load balancing is suggested as an example rationale for this, but it is not the only interpretation of multiple NHLFE entries. Other interpretations may exist and the potential interaction with load balancing should be considered. Multiple NHLFEs may be present to represent:

(a) the Incoming FEC / label set is to be multicast
(b) when route selection based on the EXP field in addition to the label is required.

If both multicast and load balancing functions are required, it is necessary to disambiguate the scope of the operations. The way we have defined the load balancing operation, it partitions a set of input traffic (defined as FECs or Labels) across a set of output LSPs. One (or more) of the arriving FECs may be multicast to both the set of load balanced LSPs as well as other LSPs. This would imply that the packet replication (multicast) function occurs before
the load balancing.

When the route selection is based on the EXP field, this appears to simply be a special case of the policy based load-balancing approach. It is suggested replicating NHLFEs for this purpose be deprecated and the more generic policy based approach be used to specify an FEC/label space partition based on the EXP field.

In this draft, we take the perspective that the load balancing function is more properly considered as part of the classification function. This allows us to preserve a mapping of an FEC into one NHLFE for unicast. While classification of incoming flows into FECs is commonly thought of as an operation on some tuple of packet headers, this is not the only basis for classification - router state can also be used. For example, the source port of a flow may be a useful basis on which to discriminate flows. Equally, a "random number" generated within the router may be attractive as the basis for allocating flows for a load balancing objective. Some algorithm within the router, which may include some hash function on the packet headers, may generate the "random number".

6. MPLS Policies for Load Balancing

MPLS load balancing partitions an incoming stream of traffic across multiple LSPs. The load balancing policy, as well as the ingress LSR where the policy is enforced, must be able to distinctly identify LSPs. We do not, in this draft, address the issue of how LSPs are established. It is assumed, however, that the PDP that installs the load balancing policy, has knowledge of the existing LSPs and is able to identify them in policy rules. One way to achieve this is through the binding of a label to an LSP.

An example MPLS load-balancing policy may state, for the simple case of balancing across two LSPs, "If traffic matches classifier C, then forward on LSP L1, else forward on LSP L2." Classification can be done on a number of parameters, such as packet header fields, incoming labels, etc. The classification conditions of an MPLS load-balancing policy are thus effectively constrained to be able to specify the FEC in terms that can be resolved into MPLS packet classification MIB parameters.

Forwarding traffic on an LSP can be achieved by tagging the traffic with the appropriate label corresponding to the LSP. MPLS load-balancing policy actions typically result in the definition of a new aggregate FEC to be forwarded down a specific LSP. This would typically be achieved by appropriate provisioning of the FEC and routing tables (FTN and ILM).

The basis for partitioning the traffic can be static or dynamic. Dynamic load balancing can be based on a dynamic administrative control (e.g. Time of Day), or it can form a closed control loop with some measured network parameter.
Static Partitioning of the Load can be based on information carried within the packet header (e.g. source / destination addresses, Source / destination port numbers, packet size, protocol ID etc.). Static partitioning can also be based on other information available at the LSR (e.g. the arriving physical interface). However if load partition is truly static, or at least very slowly changing (say < 1 change / day ?) then the need for a policy based control of this provisioning information maybe debatable and a direct manipulation of the LSR MIB may suffice.

A control-loop based load-balancing scheme seeks to balance the load close to some objective, subject to error in the measurements and delays in the feedback loop etc. The objective may be based on a fraction of the input traffic to be sent down a link (e.g. 20% down LSP (abd) and 80% down LSP (acd)) in which case some measurement of the input traffic is required. The objective may also be based on avoiding congestive loss in which case some loss metric is required.

The metrics required for control loop load balancing may be derived from information available locally at the upstream LSR, or may be triggered by events distributed elsewhere in the network. In the latter case, the metrics must be delivered to the Policy Decision Point. Obviously, locally derived trigger conditions would be expected to avoid the propagation delays etc. associated with the general distributed case. Frequent notification of the state of these metrics increases network traffic which may be undesirable. This draft does not seek to provide guidance on the appropriate rate of notification for metric updates.

Consider the case of a single large flow that must be load balanced across a set of links. In this case policies based solely on the packet headers may be inadequate and some other approach (e.g. based on a random number generated within the router) may be required. Note that sequence integrity of the aggregate FEC forwarded over a set of load balancing LSPs may not be preserved under such a regime.

7. Security Considerations

The policy system and the MPLS system both have their inherent security issues, which this document does not attempt to resolve.

The Policy system provides a mechanism to configure the LSPs within LSRs. Any thing that can be configured can also be incorrectly configured with potentially disastrous results. The policy system can help to secure the MPLS system by providing appropriate controls on the LSP life cycle. Conversely, if the security of the Policy system is compromised, then this may impact any MPLS systems controlled by that policy system.

Use of the COPS protocol within the policy system between the PEP/PDP allows the use of message level security for authentication, replay protection and message integrity. Existing protocols such as IPSEC or TLS can also be used to authenticate and secure the
channel. The COPS protocol also provides a reliable transport mechanism with a session keep-alive.

The MPLS network is not expected to impact the security of the Policy system.

Further security considerations of Policy-Enabled MPLS networks is for further study.

8. Conclusions / Further Work

Policy-based networking approaches provide a flexible mechanism that can be utilized to support certain traffic engineering objectives. In this draft we have illustrated the potential role for traffic engineering policy approaches to load balancing in MPLS networks. In future drafts we expect to extend this work to cover more sophisticated notions of QoS in addition to "best effort" MPLS services, and into MPLS PIB proposals.

The authors of this draft believe that portions of this material should be incorporated into working group drafts within the scope of the tewg, mpls and possibly other working groups.

9. References


10. Acknowledgments

Indra Widjaja provided several comments on an earlier draft.

11. Author’s Addresses

Steven Wright
Science & Technology
BellSouth Telecommunications
41G70 BSC
675 West Peachtree St. NE.
Atlanta, GA 30375
Phone +1 (404) 332-2194
Email: steven.wright@snt.bellsouth.com

Francis Reichmeyer
IPHighway, Inc.
55 New York Avenue
Framingham, MA 01701
Phone +1 (201) 655-8714
Email: franr@iphhighway.com

Robert Jaeger
Laboratory for Telecommunications Science,