

TEAS Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 21, 2016

Y. Zhuang
Q. Wu
Huawei
A. Farrel
Juniper Networks
March 18, 2016

Architecture for Scheduled Use of Resources
draft-zhuang-teas-scheduled-resources-01

Abstract

Time-Scheduled reservation of traffic engineering (TE) resources can be used to provide resource booking for TE Label Switched Paths so as to better guarantee services for customers and to improve the efficiency of network resource usage into the future. This document provides a framework that describes and discusses the architecture for the scheduled reservation of TE resources. This document does not describe specific protocols or protocol extensions needed to realize this service.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on March 21, 2016.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Problem statement	3
2.1.	Provisioning TE-LSPs and TE Resources	3
2.2.	Selecting the Path of an LSP	3
2.3.	Looking at Future Demands on TE Resources	4
2.4.	Requisite State Information	4
3.	Architectural Concepts	5
3.1.	Where is Scheduling State Held?	5
3.2.	What State is Held?	7
4.	Architecture Overview	7
4.1.	Service Request	8
4.2.	Initialization and Recovery	9
5.	Acknowledgements	9
6.	Informative References	10
	Authors' Addresses	10

[1.](#) Introduction

Traffic Engineering Label Switched Paths (TE-LSPs) are connection oriented tunnels in packet and non-packet networks [[RFC3209](#)], [[RFC3945](#)]. TE-LSPs may reserve network resources for use by the traffic they carry, thus providing some guarantees of service delivery and allowing a network operator to plan the use of the resources across the whole network.

In some technologies (such as wavelength switched optical networks) the resource is synonymous with the label that is switched on the path of the LSP so that it is not possible to establish an LSP that can carry traffic without assigning a concrete resource to the LSP. In other technologies (such as packet switched networks) the resources assigned to an LSP are a measure of the capacity of a link that is dedicated for use by the traffic on the LSP. In all cases, network planning consists of selecting paths for LSPs through the network so that there will be no contention for resources; LSP establishment is the act of setting up an LSP and reserving resources within the network; and network optimization or re-optimization is the process of re-positioning LSPs in the network to make the unreserved network resources more useful for potential future LSPs while ensuring that the established LSPs continue to fulfil their objectives.

It is often the case that it is known that an LSP will be needed at some time in the future. While a path for that LSP could be computed using knowledge of the currently established LSPs and the currently available resources, this does give any degree of certainty that the necessary resources will be available when it is time to set up the new LSP. Yet setting up the LSP ahead of the time when it is needed (which would guarantee the availability of the resources) is wasteful since the network resources could be used for some other purpose in the meantime.

Similarly, it may be known that an LSP will no longer be needed after some future time and that it will be torn down releasing the network resources that were assigned to it. This information can be helpful in planning how a future LSP is placed in the network.

Time-Scheduled (TS) reservation of TE resources can be used to provide resource booking for TE-LSPs so as to better guarantee services for customers and to improve the efficiency of network resource usage into the future. This document provides a framework that describes and discusses the architecture for the scheduled reservation of TE resources. This document does not describe specific protocols or protocol extensions needed to realize this service.

2. Problem statement

2.1. Provisioning TE-LSPs and TE Resources

TE-LSPs in existing networks are provisioned using RSVP-TE as a signaling protocol [[RFC3209](#)][RFC3473] or by direct control of network elements such as in the Software Defined Networking paradigm.

TE resources are reserved at the point of use. That is, the resources (wavelengths, timeslots, bandwidth, etc.) are reserved for use on a specific link and are tracked by the Label Switching Routers (LSRs) at the end points of the link. Those LSRs learn which resources to reserve during the LSP setup process.

The use of TE resources can be varied by changing the parameters of the LSP that uses them, and the resources can be released by tearing down the LSP.

2.2. Selecting the Path of an LSP

Although TE-LSPs can determine their paths hop-by-hop using the shortest path toward the destination to route the signaling protocol messages [[RFC3209](#)], in practice this option is not applied because it does not look far enough ahead into the network to verify that the

desired resources are available. Instead, the full length of the path of an LSP is computed ahead of time either by the head-end LSR of a signaled LSP, or by Path Computation Element (PCE) functionality in a dedicated server or built into network management software [[RFC4655](#)].

Such full-path computation is applied in order that an end-to-end view of the available resources in the network can be used to determine the best likelihood of establishing a viable LSP that meets the service requirements. Even in this situation, however, it is possible that two LSPs being set up at the same time will compete for scarce network resources meaning that one or both of them will fail to be established. This situation is avoided by using a centralized PCE that is aware of the LSP setup requests that are in progress.

[2.3.](#) Looking at Future Demands on TE Resources

While path computation as described in [Section 3.2](#) takes account of the currently available network resources, and can act to place LSPs in the network so that there is the best possibility of future LSPs being accommodated, it cannot handle all eventualities. It is simple to construct scenarios where LSPs that are placed one at a time lead to further LSPs being blocked, but where foreknowledge of all of the LSPs would have made it possible for them all to be set up.

If, therefore, we were able to know in advance what LSPs were going to be requested we could plan for them and ensure resources were available. Furthermore, such an approach enables a commitment to be made to a service user that an LSP will be set up and available at a specific time.

This service can be achieved by tracking the current use of network resources and also a future view of the resource usage. We call this time-scheduled TE (TS-TE) resource reservation.

[2.4.](#) Requisite State Information

In order to achieve the TS-TE resource reservation, the use of resources on the path needs to be scheduled. Scheduling state is used to indicate when resources are reserved and when they are available for use.

A simple information model for one piece of scheduling state is as follows:


```
{ link id;  
  resource id or reserved capacity;  
  reservation start time;  
  reservation end time  
}
```

The resource that is scheduled can be link capacity, physical resources on a link, CPU utilization, memory, buffers on an interfaces, etc. For any one resource there could be multiple pieces of scheduling state, and for any one link, the timing windows might overlap.

There are multiple ways to realize this information model and different ways to store the data. For example, the resource state could be expressed as a start time and a duration. This document does not spend any more time on discussion of encoding of state information except to discuss the location of storage of the state information and the recovery of the information after failure events.

This scheduling state information can be used by applications to book resources for future or now, so as to maximize chance of services being delivered. Also, it can avoid contention for resources of LSPs.

3. Architectural Concepts

This section examines several important architectural concepts that lead to design decisions that will influence how networks can achieve TS-TE in a scalable and robust manner.

3.1. Where is Scheduling State Held?

The scheduling state information described in [Section 3.4](#) has to be held somewhere. There are two places where this makes sense:

- o In the network nodes where the resources exist;
- o In a central scheduling controller where decisions about resource allocation are made.

The first of these makes policing of resource allocation easier. It means that many points in the network can request immediate or scheduled LSPs with the associated resource reservation and that all such requests can be correlated at the point where the resources are allocated. However, this approach has some scaling and technical problems:

- o The most obvious issue is that each network node must retain the full time-based state for all of its resources. In a busy network with a high arrival rate of new LSPs and a low hold time for each LSP, this could be a lot of state. Yet network nodes are normally implemented with minimal spare memory.
- o In order that path computation can be performed, the computing entity normally known as a Path Computation Element (PCE) [[RFC4655](#)] needs access to a database of available links and nodes in the network, and of the TE properties of the links. This database is known as the Traffic Engineering Database (TED) and is usually populated from information advertised in the IGP by each of the network nodes or exported using BGP-LS [I-D.ietf-idr-ls-distribution]. To be able to compute a path for a future LSP the PCE needs to populate the TED with all of the future resource availability: if this information is held on the network nodes it must also be advertised in the IGP. This could be a significant scaling issue for the IGP and the network nodes as all of the advertised information is held at every network node and must be periodically refreshed by the IGP.
- o When a normal node restarts it can recover resource reservation state from the forwarding hardware, from NVRAM, or from adjacent nodes through the signaling protocol [[RFC5063](#)]. If scheduling state is held at the network nodes it must also be recovered after the restart of a network node. This cannot be achieved from the forwarding hardware because the reservation will not have been made, could require additional expensive NVRAM, or might require that all adjacent nodes also have the scheduling state in order to reinstall it on the restarting node. This is potentially complex processing with scaling and cost implications.

Conversely, if the scheduling state is held centrally it is easily available at the point of use. That is, the PCE can utilize the state to plan future LSPs and can update that stored information with the scheduled reservation of resources for those future LSPs. This approach also has several issues:

- * If there are multiple controllers then they must synchronise their stored scheduling state as they each plan future LSPs, and must have a mechanism to resolve resource contention. This is relatively simple and is mitigated by the fact that there is ample processing time to replan future LSPs in the case of resource contention.
- * If other sources of immediate LSPs are allowed (for example, other controllers or autonomous action by head-end LSRs) then the changes in resource availability caused by the setup or

teardown of these LSPs must be reflected in the TED (by use of the IGP as currently) and may have an impact of planned future LSPs. This impact can be mitigated by replanning future LSPs or through LSP preemption.

- * If other sources of planned LSPs are allowed, they can request path computation and resource reservation from the centralized PCE using the PCE Communication Protocol (PCEP) [[RFC5440](#)].
- * If the scheduling state is held centrally at a PCE, the state must be held and restored after a system restart. This is relatively easy to achieve on a central server that can have access to non-volatile storage.

Thus the architectural conclusion is that scheduling state should be held centrally at the point of use and not in the network devices.

3.2. What State is Held?

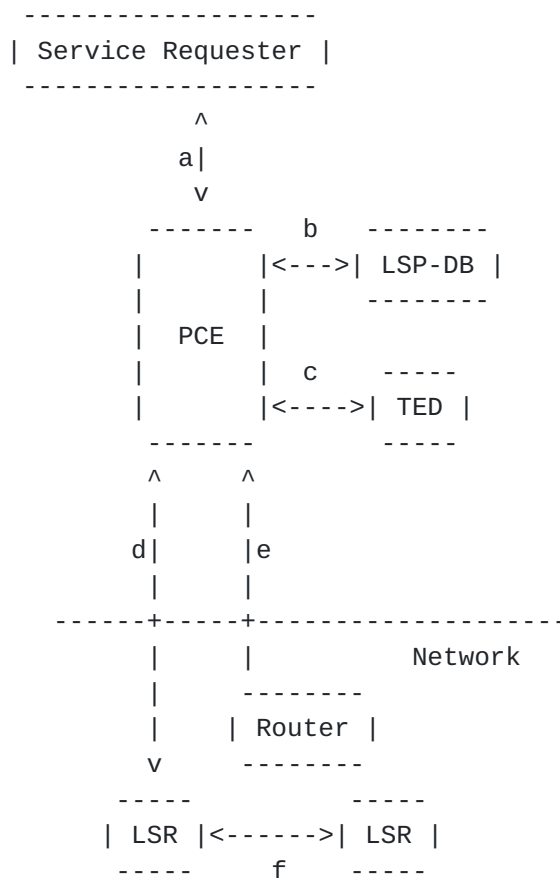
As already described, the PCE needs access to an enhanced, time-based TED. But it must be noted that service requests for future LSPs are known in terms of the LSPs whose paths are computed and for which resources are scheduled. For example, if the requester of a future LSP decides to cancel the request or to modify the request, the PCE must be able to map this to the resources that were reserved.

This suggests that the PCE needs an LSP Database {LSP-DB} [I-D.ietf-pce-stateful-pce] that contains information not only about LSPs that are active in the network, but also those that are planned.

It is an implementation choice how the TED and LSP-DB are stored both for dynamic use and for recovery after failure or restart, but it may be noted that all of the information in the scheduled TED can be recovered from the active network state and from the scheduled LSP-DB.

4. Architecture Overview

The architectural considerations and conclusions described in the previous section lead to the architecture described in this section.



Reference Architecture for Scheduled Use of Resources

4.1. Service Request

As shown in Figure 1, some component in the network requests a service. This may be an application, a Network Management System (NMS), an LSR, or any component that qualifies as a Path Computation Client (PCC). We show this on the figure as the "Service Requester" and it sends a request to the PCE for an LSP to be set up at some time (either now or in the future). The request, indicated on Figure 1 by the arrow (a) includes all of the parameters of the LSP that the requester wishes to supply such as bandwidth, start time, and end time.

The PCE enters the LSP request in its LSP-DB (b), and uses information from its TED (c) to compute a path. It updates the future resource availability in the TED so that further path computations can take account of the scheduled resource usage.

When it is time for the LSP to be set up, the PCE sends a PCEP Initiate request to the head end LSR (d) providing the path to be

signaled as well as other parameters such as the bandwidth of the LSP.

As the LSP is signaled between LSRs (f) the use of resources in the network is updated and distributed using the IGP. This information is shared with the PCE either through the IGP or using BGP-LS (e), and the PCE updates the information stored in its TED (c).

When an LSP is no longer required (either because the Service Requester has cancelled the request, or because the LSP's scheduled lifetime has expired) the PCE can remove it. If the LSP is currently active, the PCE instructs the head-end LSR to tear it down (d), and the network resource usage will be updated by the IGP and advertised back to the PCE through the IGP or BGP-LS (e). Once the LSP is no longer active, the PCE can remove it from the LSP-DB (b).

4.2. Initialization and Recovery

When a PCE in the architecture shown in Figure 1 is initialized, it must learn state from the network, from its stored databases, and potentially from other PCEs in the network.

The first step is to get an accurate view of the topology and resource availability in the network. This would normally involve reading the state direct from the network via the IGP or BGP-LS (e), but might include receiving a copy of the TED from another PCE. Note that a TED stored from a previous instantiation of the PCE is unlikely to be valid.

Next, the PCE must construct a time-based TED to show scheduled resource usage. How it does this is implementation specific: it may recover a time-based TED previously saved to non-volatile storage, or it may reconstruct the time-based TED from information retrieved from the LSP-DB previously saved to non-volatile storage. If there is more than one PCE active in the network, the recovering PCE will need to synchronize the LSP-DB and time-based TED with other PCEs.

5. Acknowledgements

This work has benefited from the discussions of resource scheduling over the years. In particular the DRAGON project [The-DRAGON-Project] and [[draft-yong-ccamp-ason-gmpls-autobw-service-00](#)] both of which provide approaches to auto-bandwidth services in GMPLS networks.

6. Informative References

[DRAGON] "http://www.maxgigapop.net/wp-content/uploads/
The-DRAGON-Project.pdf".

[I-D.ietf-pce-stateful-pce]
Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP
Extensions for Stateful PCE", [draft-ietf-pce-stateful-
pce-11](#) (work in progress), April 2015.

Authors' Addresses

Yan Zhuang
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: zhuangyan.zhuang@huawei.com

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: bill.wu@huawei.com

Adrian Farrel
Juniper Networks

Email: adrian@olddog.co.uk

