Chapter 1

Traffic Engineering Overview

The task of mapping traffic flows onto an existing physical topology is called *traffic engineering* . Traffic engineering provides the ability to move traffic flow away from the shortest path selected by the Interior Gateway Protocol (IGP) and onto a potentially less congested physical path across a network.

Traffic engineering provides the capabilities to:

- Route primary paths around known bottlenecks or points of congestion in the network.
- Provide precise control over how traffic is rerouted when the primary path is faced with single or multiple failures.
- Provide more efficient use of available aggregate bandwidth and long-haul fiber by ensuring that subsets of the network do not become overutilized while other subsets of the network along potential alternate paths are underutilized.
- Maximize operational efficiency.
- Enhance the traffic-oriented performance characteristics of the network by minimizing packet loss, minimizing prolonged periods of congestion, and maximizing throughput.
- Enhance statistically bound performance characteristics of the network (such as loss ratio, delay variation, and transfer delay) required to support a multiservices Internet.

This chapter discusses the following topics:

- Components of Traffic Engineering on page 4
- Flexible LSP Calculation and Configuration on page 7

Components of Traffic Engineering

In the JUNOS software, traffic engineering is implemented with Multiprotocol Label Switching (MPLS) and the Resource Reservation Protocol (RSVP). Traffic engineering is composed of four functional components:

Packet Forwarding Component on page 4

Information Distribution Component on page 5

Path Selection Component on page 5

Signaling Component on page 7

Packet Forwarding Component

The packet forwarding component of the JUNOS traffic engineering architecture is MPLS, which is responsible for directing a flow of IP packets along a predetermined path across a network. This path is called a *label-switched path (LSP)* . LSPs are simplex; that is, the traffic flows in one direction from the head-end (ingress) router to a tail-end (egress) router. Duplex traffic requires two LSPs: one LSP to carry traffic in each direction. An LSP is created by the concatenation of one or more label-switched hops, allowing a packet to be forwarded from one router to another across the MPLS domain.

When an ingress router receives an IP packet, it adds an MPLS header to the packet and forwards it to the next router in the LSP. The labeled packet is forwarded along the LSP by each router until it reaches the tail end of the LSP, at which point the MPLS header is removed and the packet is forwarded based on Layer 3 information such as the IP destination address. The key purpose in this scheme is that the physical path of the LSP not be limited to what the IGP would choose as the shortest path to reach the destination IP address.

This section discusses the following topics:

Packet Forwarding Based on Label Swapping on page 4

Example of How a Packet Traverses an MPLS Backbone on page 5

Packet Forwarding Based on Label Swapping

The packet forwarding process at each router is based on the concept of *label swapping*. This concept is similar to what occurs at each ATM switch in a PVC. Each MPLS packet carries a 4-byte encapsulation header that contains a 20-bit, fixed-length label field. When a packet containing a label arrives at a router, the router examines the label and uses it as an index into its MPLS forwarding table. Each entry in the forwarding table contains an interface-inbound label pair mapped to a set of forwarding information that is applied to all packets arriving on the specific interface with the same inbound label.

Example of How a Packet Traverses an MPLS Backbone

This section describes how an IP packet is processed as it traverses an MPLS backbone network.

At the entry edge of the MPLS backbone, the IP header is examined by the ingress router. Based on this analysis, the packet is classified, assigned a label, encapsulated in an MPLS header, and forwarded toward the next-hop in the LSP. MPLS provides a high degree of flexibility in the way that an IP packet can be assigned to an LSP. For example, in the JUNOS traffic engineering implementation, all packets arriving at the ingress router that are destined to exit the MPLS domain at the same egress router are forwarded along the same LSP.

Once the packet begins to traverse the LSP, each router uses the label to make the forwarding decision. The MPLS forwarding decision is made independent of the original IP header: the incoming interface and label are used as lookup keys into the MPLS forwarding table. The old label is replaced with a new label, and the packet is forwarded to the next-hop along the LSP. This process is repeated at each router in the LSP until the packet reaches the egress router.

When the packet arrives at the egress router, the label is removed and the packet exits the MPLS domain. The packet is then forwarded based on the destination IP address contained in the packet's original IP header according to the traditional shortest path calculated by the IP routing protocol.

Information Distribution Component

Traffic engineering requires detailed knowledge about the network topology as well as dynamic information about network loading. The information distribution component is implemented by defining relatively simple extensions to the IGPs so that link attributes are included as part of each router's link-state advertisement. IS-IS extensions include the definition of new Type Length Values (TLVs), while OSPF extensions are implemented with opaque link-state advertisements (LSAs). The standard flooding algorithm used by the link-state IGPs ensures that link attributes are distributed to all routers in the routing domain. Some of the traffic engineering extensions to be added to the IGP link-state advertisement include maximum link bandwidth, maximum reserved link bandwidth, current bandwidth reservation, and link coloring.

Each router maintains network link attributes and topology information in a specialized traffic engineering database (TED). The TED is used exclusively for calculating explicit paths for the placement of LSPs across the physical topology. A separate database is maintained so that the subsequent traffic engineering computation is independent of the IGP and the IGP's link-state database. Meanwhile, the IGP continues its operation without modification, performing the traditional shortest-path calculation based on information contained in the router's link-state database.

Path Selection Component

After network link attributes and topology information are flooded by the IGP and placed in the TED, each ingress router uses the TED to calculate the paths for its own set of LSPs across the routing domain. The path for each LSP can be represented by either a strict or loose explicit route. An explicit route is a preconfigured sequence of routers that should be part of the physical path of the LSP. If the ingress router specifies all the routers in the LSP, the LSP is said to be identified by a *strict* explicit route. If the ingress router specifies only some of the routers in the LSP, the LSP is described as a *loose* explicit route. Support for strict and loose explicit routes allows the path selection process to be given broad latitude whenever possible, but to be constrained when necessary.

The ingress router determines the physical path for each LSP by applying a Constrained Shortest Path First (CSPF) algorithm to the information in the TED. CSPF is a shortest-path-first algorithm that has been modified to take into account specific restrictions when calculating the shortest path across the network. Input into the CSPF algorithm includes:

Topology link-state information learned from the IGP and maintained in the TED

Attributes associated with the state of network resources (such as total link bandwidth, reserved link bandwidth, available link bandwidth, and link color) that are carried by IGP extensions and stored in the TED

Administrative attributes required to support traffic traversing the proposed LSP (such as bandwidth requirements, maximum hop count, and administrative policy requirements) that are obtained from user configuration

As CSPF considers each candidate node and link for a new LSP, it either accepts or rejects a specific path component based on resource availability or whether selecting the component violates user policy constraints. The output of the CSPF calculation is an explicit route consisting of a sequence of router addresses that provides the shortest path through the network that meets the constraints. This explicit route is then passed to the signaling component, which establishes forwarding state in the routers along the LSP.

Offline Planning and Analysis

Despite the reduced management effort resulting from online path calculation, an offline planning and analysis tool is still required to optimize traffic engineering globally. Online calculation takes resource constraints into account and calculates one LSP at a time. The challenge with this approach is that it is not deterministic. The order in which an LSP is calculated plays a critical role in determining its physical path across the network. LSPs that are calculated early in the process have more resources available to them than LSPs calculated later in the process because previously calculated LSPs consume network resources. If the order in which the LSPs are calculated is changed, the resulting set of physical paths for the LSPs also can change.

An offline planning and analysis tool simultaneously examines each link’s resource constraints and the requirements of each LSP. While the offline approach can take several hours to complete, it performs global calculations, compares the results of each calculation, and then selects the best solution for the network as a whole. The output of the offline calculation is a set of LSPs that optimizes utilization of network resources. After the offline calculation is completed, the LSPs can be established in any order because each is installed following the rules for the globally optimized solution.

Signaling Component

An LSP is not known to be workable until it is actually established by the signaling component. The signaling component, which is responsible for establishing LSP state and distributing labels, relies on a number of extensions to the Resource Reservation Protocol (RSVP):

> The Explicit Route Object allows an RSVP PATH message to traverse an explicit sequence of routers that is independent of conventional shortest-path IP routing. Recall that the explicit route can be either strict or loose.
>
> The Label Request Object permits the RSVP PATH message to request that intermediate routers provide a label binding for the LSP that it is establishing.
>
> The Label Object allows RSVP to support the distribution of labels without having to change its existing mechanisms. Because the RSVP RESV message follows the reverse path of the RSVP PATH message, the Label Object supports the distribution of labels from downstream nodes to upstream nodes.

Flexible LSP Calculation and Configuration

Traffic engineering involves mapping traffic flow onto a physical topology. You can determine the paths online using constraint-based routing. Regardless of how the physical path is calculated, the forwarding state is installed across the network using RSVP.

The JUNOS software supports a number of different ways to route and configure an LSP:

> You can calculate the full path for the LSP offline and individually configure each router in the LSP with the necessary static forwarding state. This is analogous to how some ISPs currently configure their IP-over-ATM cores.
>
> You can calculate the full path for the LSP offline and statically configure the ingress router with the full path. The ingress router then uses RSVP as a dynamic signaling protocol to install a forwarding state in each router along the LSP.
>
> You can rely on constraint-based routing to perform dynamic online LSP calculation. You configure the constraints for each LSP, and then the network itself determines the path that best meets those constraints. Specifically, the ingress router calculates the entire LSP based on the constraints and then initiates signaling across the network.
>
> You can calculate a partial path for an LSP offline and statically configure the ingress router with a subset of the routers in the path and then permit online calculation to determine the complete path.
>
> For example, consider a topology that includes two east-west paths across the United States: one in the north through Chicago and one in the south through Dallas. If you want to establish an LSP between a router in New York and one in San Francisco, you can configure the partial path for the LSP to include a single loose-routed hop of a router in Dallas. The result is an LSP routed along the southern path. The ingress router uses CSPF to compute the complete path and uses RSVP to install the forwarding state along the LSP.

> You can configure the ingress router with no constraints whatsoever. In this case, normal IGP shortest-path routing is used to determine the path of the LSP. This configuration does not provide any value in terms of traffic engineering. However, it is easy and might be useful in situations when services such as Virtual Private Networks (VPNs) are needed.

In all these cases, you can specify any number of LSPs as backups for the primary LSP, thus allowing you to combine more than one configuration approach. For example, you might explicitly compute the primary path offline, set the secondary path to be constraint-based, and have the tertiary path be unconstrained. If a circuit on which the primary LSP is routed fails, the ingress router notices the outage from error notifications received from a downstream router or by the expiration of RSVP soft-state information. Then, the router dynamically forwards traffic to a hot-standby LSP or calls on RSVP to create a forwarding state for a new backup LSP.