Path Computation Element Protocol

This chapter provides a comprehensive overview of the Path Computation Element Protocol (PCEP), a mechanism for centralized path computation and orchestration in Segment Routing networks.

It explains the roles of Path Computation Elements (PCEs) and Path Computation Clients (PCCs), detailing how they interact to request, compute, and manage Segment Routing over IPv6 Label Switched Paths (LSPs). The chapter also covers PCEP's benefits, session establishment guidelines, authentication methods, various timers for session stability, and the PCC-centric redundancy model for high availability.

Path computation element protocol

The path computation element protocol (PCEP) is a set of procedures that

- enables Path Computation Clients (PCCs) to report and delegate control of head-end SR paths to Path Computation Element (PCE) peers
- allows PCEs to request updates or modifications to path parameters, and
- supports stateful models where PCEs can initiate computations for network-wide orchestration and establishes channels over TCP with a lightweight keep-alive (KA) mechanism.

Key concepts for PCEP

- Path Computation Element (PCE): PCE is a network element that computes and orchestrates paths in a segment routing network. It identifies each segment using an IPv6 address known as a Segment Identifier or SID rather than an MPLS label, and can optimize, update, and orchestrate SRv6 paths across the network.
- Path Computation Client (PCC): PCC is a network device that interacts with the PCE to request path computations, delegate path control, and report the status of SRv6 LSPs. The PCC programs the IPv6 SIDs into the Segment Routing Header, or SRH, of packets.
- Label Switched Path (LSP): LSP is a sequence of IPv6 SIDs that define the forwarding path in the network. These SIDs are carried in the SRH. The path is established and managed.

- Maximum SID Depth (MSD): MSD defines the maximum number of Segment Identifiers (SIDs) that can be included in a Segment Routing (SR) path. It acts as a critical constraint during path computation, ensuring that computed paths adhere to the capabilities of network devices or specific policy requirements. PCCs can signal their MSD capabilities or requirements to PCEs via PCEP. This influences how paths are calculated and validated.
- PCEP related timers: PCEP timers are configurable parameters that govern the operational aspects of PCEP sessions, including session liveness (keepalives) and the management of delegated Segment Routing policies. They ensure session stability and proper handling of policy states, especially during network events or PCE unreachability. For more information, see PCEP related Timers.
- PCEP Authentication: PCEP Authentication is the security mechanisms used to verify the identity of PCEP peers and ensure the integrity of the communication channel. It prevents unauthorized entities from participating in or disrupting PCEP sessions. For more information, see PCEP authentication, on page 7.
- PCC-Centric Redundancy: A high-availability model for PCEP that centralizes LSP delegation control at the PCC. This model ensures continuous operation and efficient failover/failback of LSPs by managing re-delegation to alternate PCEs when the primary PCE becomes unavailable, and then facilitating a return to the original PCE upon its recovery. For more information, see PCC-Centric redundancy, on page 10.

Benefits of PCEP

PCEP provides significant advantages for managing and optimizing network paths. These benefits enhance network efficiency, flexibility, and resilience, and include:

- Centralized path optimization: Enables the use of a centralized PCE to compute optimal network paths based on global network knowledge, rather than relying on distributed algorithms at each router.
- Dynamic traffic engineering: Supports real-time adjustment of SRv6 paths in response to changing network conditions, policies, or failures, improving resource utilization and network resilience.
- Stateful control and delegation: Allows stateful PCEs to maintain information about existing paths and take full or partial control of path setup and modification, leading to more intelligent and coordinated path management.
- Simplified operations: Reduces manual configuration and complexity by enabling automation of path computation, provisioning, and optimization through a standardized protocol.
- Support for advanced constraints: Facilitates computation of paths based on diverse constraints, such as bandwidth, latency, disjointness, policy, which may not be supported in traditional distributed routing.
- Enhanced Fault Recovery: Enables rapid re-computation and rerouting of paths in case of network failures or congestion. This improves overall network availability and reliability.
- Security Features: Offers secure session establishment and authentication, such as TCP-AO, helping protect the control plane communications.

Usage Guidelines for Path Computation Element Protocol

PCEP session acceptance conditions

For a configured PCE peer to successfully establish a PCEP session with the PCE, the following conditions must be satisfied:

- The total number of PCEP sessions on the PCE must not exceed its configured limit.
- The Keepalive (KA) interval indicated by the PCC must be acceptable to the PCE.

MSD

For cases with path computation at PCE, a PCC can signal its MSD to the PCE in the following ways:

- During PCEP session establishment – The signaled MSD is treated as a node-wide property.
- During PCEP LSP path request – The signaled MSD is treated as an LSP property.
- Local SR Policy: MSD is configured using the **segment-routing traffic-eng policy** command.

Note If the configured MSD values are different, the per-LSP MSD takes precedence over the per-node MSD.

After path computation, the resulting uSID stack size is verified against the MSD requirement.

- If the uSID stack size is larger than the MSD and path computation is performed by PCE, then the PCE returns a "no path" response to the PCC.
- If the uSID stack size is larger than the MSD and path computation is performed by PCC, then the PCC will not install the path.

Note A sub-optimal path (if one exists) that satisfies the MSD constraint could be computed in the following cases:

- For a dynamic path with TE metric, when the PCE is configured with the **pce segment-routing te-latency** command or the PCC is configured with the **segment-routing traffic-eng te-latency** command.
- For a dynamic path with LATENCY metric
- For a dynamic path with affinity constraints

For example, if the PCC MSD is 4 and the optimal path (with an accumulated metric of 100) requires 5 uSIDs, but a sub-optimal path exists (with accumulated metric of 110) requiring 4 uSIDs, then the sub-optimal path is installed.

How PCEP Works

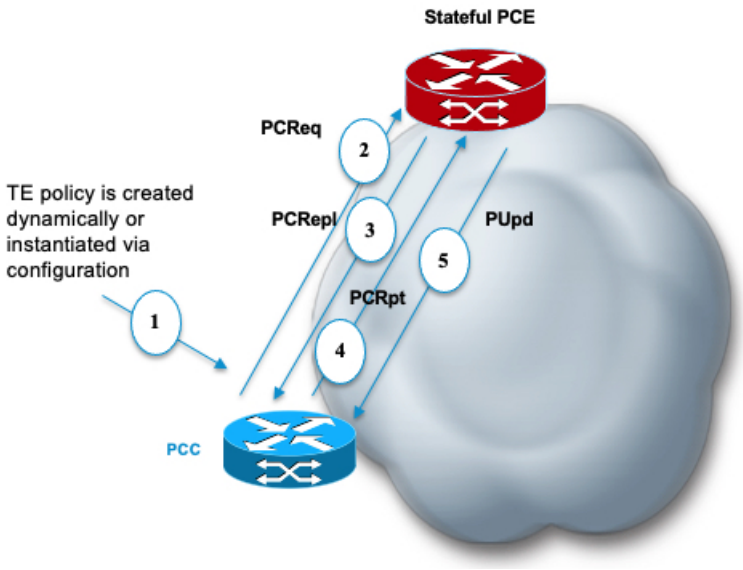
PCEP enables a PCC to request path computations from a PCE.

Summary

The key components involved in this workflow are the PCC and the PCE. The PCC initiates path computation requests and activates SR policies, while the PCE computes optimal paths and maintains policy delegation. Together, these components establish and dynamically maintain SRv6-TE policies, ensuring efficient traffic steering and network adaptability in response to changing conditions.

Workflow

Figure 1: Sample workflow with Stateful PCEP



These stages describe how a sample workflow with Stateful PCEP works:

1. The PCC is configured to instantiate an SRv6-TE policy. It sends a PCEP Path Computation Request (PCReq) to the PCE, requesting a path by specifying path attributes, optimization objectives, and constraints.
2. The PCE stores the request, computes a TE metric shortest-path, and returns the computed SID list in a PCEP Path Computation Reply (PCRepl).
3. The PCC allocates a Binding SID (BSID) and activates the SR Policy using the SID list computed by the PCE. The PCC then sends a Path Computation Report (PCRpt) to the PCE, delegating the SR Policy to the PCE and including the BSID
4. PCE updates paths as needed: The PCE updates the paths when required, for example, following a multi-domain topology change that impacts connectivity, ensuring the SR Policy remains optimized

Configure head-end router as a PCEP Path Computation Client

Before you begin

Ensure that the PCC and PCE addresses are routable to allow the TCP connection (for exchanging PCEP messages) to be established between them

Procedure

Step 1 Enable the SR-TE head-end router as a PCEP client (PCC) with 2 PCEP servers (PCE) with different precedence values.

Example:

```
Node1(config-sr)# traffic-eng
Node1(config-sr-te)# pcc
Node1(config-sr-te-pcc)# source-address ipv6 cafe:0:1::1
```

```
Node1(config-sr-te-pcc)# pce address ipv6 cafe:0:2::2
Node1(config-pcc-pce)# precedence 10
Node1(config-pcc-pce)# exit
Node1(config-sr-te-pcc)# pce address ipv6 cafe:0:3::3
Node1(config-pcc-pce)# precedence 20
Node1(config-pcc-pce)# exit
```

Step 2 Enable PCEP reporting for all policies in the node.

Example:

```
Node1(config-sr-te-pcc)# report-all
Node1(config-sr-te-pcc)# exit
```

Step 3 Set the maximum SID Depth (MSD).

Example:

```
Node1(config-sr-te)# srv6
Node1(config-sr-te-srv6)# maximum-sid-depth 5
Node1(config-sr-te-srv6)# exit
```

Step 4 Enable SR-TE related syslogs.

Example:

Node1(config-sr-te)# **logging** Node1(config-sr-te-log)# **policy status** Node1(config-sr-te-log)# **exit** Node1(config-sr-te)#

Step 5 Use the **show running-config** command to review the current configuration.

Example:

```
segment-routing
 traffic-eng
  srv6
   maximum-sid-depth 5
  !
  logging
   policy status
  !
  pcc
   source-address ipv6 cafe:0:1::1
   pce address ipv6 cafe:0:2::2
    precedence 10
   !
   pce address ipv6 cafe:0:3::3
    precedence 20
   !
   report-all
  !
 !
!
```

Step 6 Verify the status and summary of IPv6 PCEP peers for SRv6-TE) on the router.

Example:

```
Node1# show segment-routing traffic-eng pcc ipv6 peer brief

              Address   Precedence        State    Learned From
-------------------- ------------ ------------ ---------------
         cafe:0:2::2           10           up          config
         cafe:0:3::3           20           up          config
```

```
Node1# show segment-routing traffic-eng pcc ipv6 peer detail

PCC's peer database:
--------------------

Peer address: cafe:0:2::2
  Precedence: 10, (best PCE)
  State up
  Capabilities: Stateful, Update, Segment-Routing, Instantiation
  PCEP has been up for: 01:22:23
  Local keepalive timer is 30 seconds
  Remote keepalive timer is 30 seconds
  Local dead timer is 120 seconds
  Remote dead timer is 120 seconds
  Authentication: None
  Statistics:
    Open messages:      rx 1        |  tx 1
    Close messages:     rx 0        |  tx 0
    Keepalive messages: rx 164      |  tx 163
    Error messages:     rx 0        |  tx 0
    Report messages:    rx 0        |  tx 110
    Update messages:    rx 36       |  tx 0

Peer address: cafe:0:3::3
  Precedence: 20
  State up
  Capabilities: Stateful, Update, Segment-Routing, Instantiation
  PCEP has been up for: 01:21:48
  Local keepalive timer is 30 seconds
  Remote keepalive timer is 30 seconds
  Local dead timer is 120 seconds
  Remote dead timer is 120 seconds
  Authentication: None
  Statistics:
    Open messages:      rx 1        |  tx 1
    Close messages:     rx 0        |  tx 0
    Keepalive messages: rx 164      |  tx 162
    Error messages:     rx 0        |  tx 0
    Report messages:    rx 0        |  tx 82
    Update messages:    rx 0        |  tx 0
```

Step 7 (optional) Enable ECMP-aware path computation for TE metric to customize the SR-TE path calculation.

Example:

```
Router(config-sr-te)# te-latency
```

Example:

Note
ECMP-aware path computation is enabled by default for IGP and LATENCY metrics.

What to do next

- (optinal) Configure PCEP Authentication
- (optinal) Configure PCEP-Related Timers

- (optinal) Configure PCEP Redundancy Type

PCEP authentication

The PCEP session authentication is a security mechanism that

- establishes a secure and trusted communication channel between a PCC and a PCE
- ensures that only authorized devices can participate in PCEP sessions, and
- protects against unauthorized access and data tampering.

Methods of PCEP authentication

PCEP authentication can be achieved using one of two primary methods:

- TCP Message Digest 5 (MD5) Authentication: This method uses a clear text or encrypted password for authentication. Segments lacking a Message Authentication Code (MAC) that matches the configured password are rejected.
- TCP Authentication Option (TCP-AO): TCP-AO uses Message Authentication Codes (MACs), which provides these benefits:
  - Protection against replays for long-lived TCP connections
  - More details on the security association with TCP connections than TCP MD5
  - A larger set of MACs with minimal system and operational changes.

  TCP-AO is compatible with Master Key Tuple (MKT) configuration, protecting connections by deriving traffic keys from the MKT and coordinating changes between endpoints. Segments lacking a MAC that matches the configured key chain are rejected.

Note TCP-AO and TCP MD5 are never permitted to be used simultaneously. TCP-AO supports IPv6, and is fully compatible with the proposed requirements for the replacement of TCP MD5.

Configure PCEP Authentication

Procedure

Step 1 Configure PCEP Authentication using using either TCP MD5 or TCP-AO.

- Configure TCP Message Digest 5 (MD5) Authentication. Specify if the password is encrypted or clear text.

```
Router(config-sr-te-pcc)# pce address ipv6 ipv6-PCE-address[password {clear | encrypted} LINE]
```

- Configure TCP Authentication Option (TCP-AO). Use the **include-tcp-options** keyword to include other TCP options in the header for MAC calculation.

```
Router(config-sr-te-pcc)# pce address ipv6 ipv6-PCE-address tcp-ao key-chain [include-tcp-options]
```

Step 2 Verify the PCEP Authentication

PCEP-Related Timers

PCEP related timers are configurable parameters that

- govern the behavior and stability of PCEP sessions and the management of delegated SR policies
- ensure proper session maintenance, and
- manage the lifecycle of SR policies initiated or delegated by a PCE.

Types of PCEP-Related Timers:

- Keepalive timer: This timer specifies the frequency (in seconds) at which keepalive messages are sent from the PCC to its PCE peers. These messages confirm the liveness of the PCEP session. Default: 30 seconds
- Deadtimer: This timer defines how long (in seconds) remote PCE peers will wait before declaring a PCEP session down if no PCEP messages are received from the PCC. It acts as a session timeout. Default: 120 seconds
- Delegation timeout: This timer determines the maximum duration (in seconds) a delegated SR policy can remain active on the PCC without an active connection to its delegating PCE. Default: 60 seconds.
- Initiated orphans timer: This timer specifies the amount of time (in seconds) a PCE-initiated SR policy will remain delegated to a PCE peer that has become unreachable by the PCC. It allows for a grace period for the PCE to reconnect. Default: 180 seconds
- Initiated state timer: This timer defines the duration (in seconds) a PCE-initiated SR policy will remain programmed and active in forwarding on the head-end, even when it is not currently delegated to any PCE (e.g., after the orphan timer expires). Default: 600 seconds.

How PCE-Initiated SR Policy Timers Operate

Summary

The key components involved in this process are:

- Path Computation Element (PCE): The entity that initiates and delegates SR policies, for example, PCE A.
- Segment Routing Policy: The network path definition managed by the PCE, for example, Policy P.
- Head-end Router (PCC): The device that receives, programs, and manages the delegated SR policy, for example, Head-end N.
- PCEP Timers: Specifically, the `initiated orphans` and `initiated state` timers that govern the policy's lifecycle during PCE unreachability.

The process involves the head-end router using initiated orphans and initiated state timers to either preserve the SR policy through re-delegation or remove it from forwarding if no PCE takes ownership within defined timeframes.

Workflow

These are the stages for PCE-initiated SR policy timers:

1. Policy Instantiation and Initial Delegation: PCE A instantiates SR Policy P at Head-end N. Head-end N then delegates Policy P to PCE A and programs it into its forwarding plane.
2. PCE Unreachability Detection: If Head-end N detects that PCE A is no longer reachable, head-end N starts both the initiated orphans and initiated state timers for SR Policy P.
3. Re-delegation Grace Period: If PCE A reconnects before the orphan timer expires, then SR policy P is automatically delegated back to its original PCE (PCE A). After the orphan timer expires, SR policy P will be eligible for delegation to any other surviving PCE(s).
4. Policy Removal: If SR policy P is not delegated to another PCE before the state timer expires, then head-end N will remove SR policy P from its forwarding.

Configure PCEP-related timers

Set timer intervals that govern PCEP session behavior and management of PCE-initiated SR policies.

Adjusting these timers allows tailored session management for specific network performance needs. You can configure the timers independently and in any order.

Procedure

Step 1 Use the **timers keepalive** command to specify how often keepalive messages are sent from the PCC to its peers.

The range is from 0 to 255 seconds; the default value is 30.

Example:

```
Router(config-sr-te-pcc)# timers keepalive 30
```

Step 2 Use the **timers deadtimer** command topecify how long the remote peers wait before bringing down the PCEP session if no PCEP messages are received from this PCC.

The range is from 1 to 255 seconds; the default value is 120.

Example:

```
Router(config-sr-te-pcc)# timers deadtimer 120
```

Step 3 Use the **timers delegation-timeout** command to specify how long a delegated SR policy can remain active without an active connection to its delegating PCE.

The range is from 0 to 3600 seconds; the default value is 60.

Example:

```
Router(config-sr-te-pcc)# timers delegation-timeout 60
```

Step 4 Use the **timers initiated orphans** command to specify the amount of time that a PCE-initiated SR policy will remain delegated to a PCE peer that is no longer reachable by the PCC,

The range is from 10 to 180 seconds; the default value is 180.

Example:

```
Router(config-sr-te-pcc)# timers initiated orphans 180
```

Step 5 Use the **timers initiated state** command to specify the amount of time that a PCE-initiated SR policy will remain programmed while not being delegated to any PCE.

The range is from 15 to 14440 seconds (24 hours); the default value is 600.

Example:

```
Router(config-sr-te-pcc)# timers initiated state 600
```

PCC-Centric redundancy

The PCC-Centric Redundancy is a high-availability model for PCEP that

- centralizes LSP delegation control at the PCC
- ensures continuous operation and efficient failover/failback of LSPs, and
- manages re-delegation to alternate PCEs when the primary PCE becomes unavailable, and then facilitates a return to the original PCE upon its recovery.

How the PCC-Centric Redundancy Model Works

Summary

The key components involved in this process are:

- PCC: The device managing the LSP delegation.
- PCE: The entity that computes and delegates LSPs.
- Label switched path (LSP): The network path being delegated and managed.
- Delegation fallback timer: A specific timer that governs the return of an LSP to its original PCE.

The PCC-centric redundancy model maintains high availability for Label Switched Paths (LSPs) in PCEP environments by managing the delegation of LSPs between primary and alternate Path Computation Elements (PCEs). The process includes automatic failover to an alternate PCE when the original becomes unreachable and graceful restoration to the primary PCE once it recovers, coordinated by a delegation fallback timer.

Workflow

These stages describe the PCC-Centric Redundancy Model.

1. Initial LSP delegation: After an LSP is created, the PCC automatically delegates it to the specific PCE that computed the LSP.

2. PCE disconnection handling: If the original PCE (to which the LSP was delegated) becomes disconnected or unreachable, the PCC automatically re-delegates the LSP to another available PCE.
3. Original PCE reconnection and fallback: If the original PCE reconnects and becomes reachable again, a "delegation fallback timer" is initiated by the PCC. Once this timer expires, the LSP is re-delegated back to the original PCE, even if that PCE has a worse preference than the currently active PCE. This action ensures that LSPs eventually return to their primary, originating PCE.

Configure PCEP Redundancy Type

This task describes how to enable the PCC-centric high-availability model for PCEP, which modifies the default LSP delegation behavior to enhance redundancy.

Procedure

Step 1 Run the **redundancy pcc-centric** command to nable PCC-centric high-availability model.

Example:

```
Router(config-sr-te-pcc)# redundancy pcc-centric
```

Step 2 Verify the PCEP Redundancy Type.