Nokia Service Router Linux

Release 23.10

P4Runtime Guide

3HE 19843 AAAA TQZZA
Edition: 01
November 2023

Nokia is committed to diversity and inclusion. We are continuously reviewing our customer documentation and consulting with standards bodies to ensure that terminology is inclusive and aligned with the industry. Our future customer documentation will be updated accordingly.

Table of contents

1 About this guide 4
1.1 Precautionary and information messages 4
1.2 Conventions 4

2 What's new 6

3 Overview 7
3.1 SR Linux `p4rt_server` process 7

4 Supported P4Runtime RPCs 9
4.1 `StreamChannel` RPC 9
4.1.1 P4Runtime client arbitration 10
4.1.2 `PacketIn` and `PacketOut` messages 11

5 Configuring SR Linux for P4Runtime 13
5.1 Identifying interfaces to P4Runtime 13
5.1.1 Configuring a port identifier for P4Runtime 13
5.1.2 Configuring a device identifier for P4RT 14
5.2 Configuring global P4Runtime server settings 14
5.3 Configuring the P4Runtime server for a network-instance 15
5.4 Configuring the P4Runtime server for UNIX sockets 16
5.5 Disconnecting P4Runtime clients 16

1 About this guide

This document describes Nokia Service Router Linux (SR Linux) support for P4Runtime and includes procedures for configuring SR Linux to operate with P4Runtime clients.

Note:
This manual covers the current release and may also contain some content to be released in later maintenance loads. See the *SR Linux Release Notes* for information about features supported in each load.

Configuration and command outputs shown in this guide are examples only; actual displays may differ depending on supported functionality and user configuration.

1.1 Precautionary and information messages

The following are information symbols used in the documentation.

DANGER: Danger warns that the described activity or situation may result in serious personal injury or death. An electric shock hazard could exist. Before you begin work on this equipment, be aware of hazards involving electrical circuitry, be familiar with networking environments, and implement accident prevention procedures.

WARNING: Warning indicates that the described activity or situation may, or will, cause equipment damage, serious performance problems, or loss of data.

Caution: Caution indicates that the described activity or situation may reduce your component or system performance.

Note: Note provides additional operational information.

Tip: Tip provides suggestions for use or best practices.

1.2 Conventions

Nokia SR Linux documentation uses the following command conventions.

- **Bold** type indicates a command that the user must enter.
- Input and output examples are displayed in `Courier` text.
- An open right-angle bracket indicates a progression of menu choices or simple command sequence (often selected from a user interface). Example: **start** > **connect to**.
- A vertical bar (|) indicates a mutually exclusive argument.
- Square brackets ([ ]) indicate optional elements.

- Braces ({ }) indicate a required choice. When braces are contained within square brackets, they indicate a required choice within an optional element.
- *Italic* type indicates a variable.

Generic IP addresses are used in examples. Replace these with the appropriate IP addresses used in the system.

2 What's new

There have been no updates in this document since it was last released.

3 Overview

Programming Protocol-Independent Packet Processors (P4) is an open-source language for programming the data plane on networking devices. P4Runtime is an API for controlling the data plane on devices defined in a P4 program. The P4 language and P4Runtime specification are maintained at p4.org.

The SR Linux eXtensible Data Path (XDP) is not programmed in P4. However, SR Linux is packaged with a fixed P4 program that provides support for marking packets for trapping to a P4Runtime client via PacketIn messages, and transmitting packets from the P4Runtime client to an interface on the device via PacketOut messages. The following fields can be used to mark frames for extraction:

- VLAN ID
- Ethertype
- TTL

This could for example be used to redirect traceroute packets with TTL=0, TTL=1, or TTL=2 to a P4runtime client, so they can be enriched with information that is not visible to the device for the following ACL rules:

- TTL=0, IPv4 (ethertype 0x0800)
- TTL=1, IPv4 (ethertype 0x0800)
- TTL=2, IPv4 (ethertype 0x0800)
- TTL=0, IPv6 (ethertype 0x86DD)
- TTL=1, IPv6 (ethertype 0x86DD)
- TTL=2, IPv6 (ethertype 0x86DD)

Another use case is to use a free ethertype to allow the P4Runtime client to transmit and receive packets on all internal links on all devices in a network as a means of topology discovery.

To accommodate these use cases, the SR Linux runs a process `p4rt_server` that runs a gRPC server that provides the interface between P4Runtime clients and SR Linux.

3.1 SR Linux `p4rt_server` process

SR Linux supports packet input/output to P4Runtime clients through the `p4rt_server` process. The `p4rt_server` process exposes instances of P4Runtime RPCs that P4Runtime clients can connect to, with mandatory arbitration to elect a single P4Runtime client as the primary (see P4Runtime client arbitration).

Instead of running multiple processes, SR Linux runs a single `p4rt_server` process with multiple sockets. The `p4rt_server` process can expose sockets in multiple network-instances, supporting both per network-instance configuration and UNIX-socket configuration, allowing the `p4rt_server` process to run on different ports or use different authentication mechanisms within different network-instances.

The `p4rt_server` process uses TCP port 9559 by default, but this port is configurable. Communication between client and server is secured using TLS, so that P4Runtime clients are authenticated using the settings in a TLS profile.

With `authenticate-client` set to `true` in the TLS profile, new connections are mutually authenticated. Each entity validates the X.509 certificate of the remote entity to ensure that the remote entity is both known and authorized to connect to the local system. See the "Using SPIFFE for client authentication (mTLS)" section in the *SR Linux Configuration Basics Guide* for information about using SPIFFE for client authentication in TLS sessions.

The `p4rt_server` process runs as the `p4rtrpc` user. The `p4rtrpc` user is installed in the `tls` group, which allows the `p4rt_server` process to read and use certificates and keys populated via `linux_mgr`.

See Configuring SR Linux for P4Runtime for information about configuring the `p4rt_server` process.

4 Supported P4Runtime RPCs

To support the use cases described in Overview, SR Linux supports the following P4Runtime RPCs:

- `StreamChannel` RPC
  Allows for session management, client arbitration, and packet injection/extraction. See StreamChannel RPC for information about how SR Linux handles `StreamChannel` RPC messages.
- `Write` RPC
  Injects rules to select packets to extract to the slow path.
- `Read` RPC
  Provides uniformity with the `Write` RPC; that is, it gives the controller the ability to query which rules have been programmed.
- `SetForwardingPipelineConfig` RPC
  Used for pushing a P4 program from a P4Runtime controller to a target SR Linux device. The P4 program must be the same program that is distributed with SR Linux.
- `GetForwardingPipelineConfig` RPC
  Used for reading out the P4 forwarding pipeline configuration.
- `Capabilities` RPC
  Allows a P4Runtime client to discover the capabilities of a target device, including the P4Runtime API version implemented by the SR Linux `p4rt_server` process.

4.1 `StreamChannel` RPC

The P4Runtime StreamChannel RPC is used for session management, arbitration, and packet I/O. It has the following definition:

```
rpc StreamChannel(stream StreamMessageRequest) returns (stream StreamMessageResponse)
```

The StreamChannel RPC has two top-level messages, `StreamMessageRequest` and `StreamMessageResponse`.

The following is an example of the `StreamMessageRequest` message.

```
message StreamMessageRequest {
    oneof update {
        MasterArbitrationUpdate arbitration = 1;
        PacketOut packet = 2;
        DigestListAck digest_ack = 3;
        .google.protobuf.Any other = 4;
    }
}
```

The SR Linux `p4rt_server` process uses fields in the `StreamMessageRequest` messages as follows:

- `MasterArbitrationUpdate`

P4Runtime clients send `StreamMessageRequest` messages to the `p4rt_server` on the SR Linux to perform arbitration via `MasterArbitrationUpdate` messages. See P4Runtime client arbitration.

- `PacketOut`
  The P4Runtime client selected as primary transmits packets via `PacketOut` messages. See PacketIn and PacketOut messages.
- The `DigestListAck` and `.google.proto.Any` messages within the `digest_ack` and `other` fields are not supported by SR Linux.

The following is an example of the `StreamMessageResponse` message:

```
message StreamMessageResponse {
    oneof update {
        MasterArbitrationUpdate arbitration = 1;
        PacketIn packet = 2;
        DigestList digest = 3;
        IdleTimeoutNotification idle_timeout_notification = 4;
        .google.protobuf.Any other = 5;
        // Used by the server to asynchronously report errors which occur when
        // processing StreamMessageRequest messages.
        StreamError error = 6;
    }
}
```

The SR Linux `p4rt_server` process uses fields in the `StreamMessageResponse` messages as follows:

- `MasterArbitrationUpdate`
  The `p4rt_server` sends a `StreamMessageResponse` message to a P4Runtime client to respond to an arbitration request via a `MasterArbitrationUpdate` message. See P4Runtime client arbitration
- `PacketIn`
  The `p4rt_server` transmits packets to the primary P4Runtime client via `PacketIn` messages. See PacketIn and PacketOut messages.
- `StreamError`
  The `p4rt_server` transmits `StreamError` message within the `error` field based on any errors occurring. See Stream Error Reporting.
- The `DigestList`, `IdleTimeoutNotification`, `.google.proto.Any` messages within the `digest`, `idle_timeout_notification`, and `other` fields are not are not supported by SR Linux.

4.1.1 P4Runtime client arbitration

P4Runtime client arbitration refers to the process by which a single P4Runtime controller becomes the "primary", with any other controllers serving as backups. Only one P4Runtime controller can be the primary. On SR Linux, P4Runtime arbitration works as described in the P4 Runtime specification: the primary is elected based on the `election_id` within the `MasterArbitrationUpdate` message in the `StreamChannel` RPC.

The `MasterArbitrationUpdate` message is defined as follows:

```
message MasterArbitrationUpdate {
    uint64 device_id = 1;
    // The role for which the primary client is being arbitrated. For use-cases
    // where multiple roles are not needed, the controller can leave this unset,
    // implying default role and full pipeline access.
```

```
    Role role = 2;
    // The stream RPC with the highest election_id is the primary. The 'primary'
    // controller instance populates this with its latest election_id. Switch
    // populates with the highest election ID it has received from all connected
    // controllers.
    Uint128 election_id = 3;
    // Switch populates this with OK for the client that is the primary, and
    // with an error status for all other connected clients (at every primary
    // client change). The controller does not populate this field.
    .google.rpc.Status status = 4;
}

message Role {
    // Uniquely identifies this role.
    string name = 3;
    // Describes the role configuration, i.e. what operations, P4 entities,
    // behaviors, etc. are in the scope of a given role. If config is not set
    // (default case), it implies all P4 objects and control behaviors are in
    // scope, i.e. full pipeline access. The format of this message is
    // out-of-scope of P4Runtime.
    .google.protobuf.Any config = 2;
}
```

Only the primary controller is allowed to send `PacketOut` messages and receive `PacketIn` messages for a specific ASIC.

In the event a controller with the highest `election_id` disconnects, the controller with the next-highest `election_id` is automatically used as the new primary, resulting in `PacketIn` messages being forwarded to the new primary, and `p4rt_server` only accepting `PacketOut` from the new primary. If a controller does not specify an `election_id`, it is considered to be the lowest `election_id` and never becomes the primary.

4.1.2 `PacketIn` and `PacketOut` messages

`PacketIn` messages are sent by `p4rt_server` to the P4Runtime client within `StreamMessage Response` messages, and `PacketOut` messages are sent by the P4Runtime client to `p4rt_server` within `StreamMessageRequest` messages.

An ACL pushed to the device via the `Write` RPC is used to mark/extract packets for `PacketIn` handling by the `p4rt_server`.

`PacketIn` and `PacketOut` messages are defined as follows:

```
// Packet sent from the controller to the switch.
message PacketOut {
    bytes payload = 1;
    // This will be based on P4 header annotated as
    // @controller_header("packet_out").
    // At most one P4 header can have this annotation.
    repeated PacketMetadata metadata = 2;
}

// Packet sent from the switch to the controller.
message PacketIn {
    bytes payload = 1;
    // This will be based on P4 header annotated as
    // @controller_header("packet_in").
    // At most one P4 header can have this annotation.
    repeated PacketMetadata metadata = 2;
```

```
}

message PacketMetadata {
    // This refers to Metadata.id coming from P4Info ControllerPacketMetadata.
    uint32 metadata_id = 1;
    bytes value = 2;
}
```

The `PacketIn` and `PacketOut` messages require that the interface a packet is received on or transmitted out of be uniquely identified. To do this, unique identifiers are configured for SR Linux interfaces. See Identifying interfaces to P4Runtime.

5 Configuring SR Linux for P4Runtime

To configure SR Linux for P4Runtime, you perform the following configuration tasks:

- Configure a TLS profile to secure communication with P4Runtime clients. See TLS profiles in the *SR Linux Configuration Basics Guide* for information about configuring TLS profiles.
- Configure interface identifiers. To allow P4Runtime clients to reference specific interfaces in `PacketIn` and `PacketOut` messages, you configure per-interface identifiers consisting of a port ID and device ID. See Identifying interfaces to P4Runtime.
- Configure global settings for the P4Runtime server, including idle timeout, session limits, and connection rate limiting. See Configuring global P4Runtime server settings.
- Configure network-instance specific settings for the P4Runtime server. See Configuring the P4Runtime server for a network-instance.
- Configure UNIX-socket specific settings for the P4Runtime server. See Configuring the P4Runtime server for UNIX sockets.
- Configure service authorization for the `p4rt` interface type. See Service authorization in the *SR Linux Configuration Basics Guide*.

5.1 Identifying interfaces to P4Runtime

The `PacketIn` and `PacketOut` messages within the P4Runtime `StreamChannel` RPC require that the interface a packet is received on or transmitted out of be uniquely identified. To do this, you configure a unique per-interface identifier, which is a tuple consisting of the following:

- A chassis-unique port identifier (known as the `interface_id`). This identifier can be manually configured, or if it is not, the system `ifIndex` for the interface is used by default.
- A chassis-unique device identifier that indicates the specific line card and ASIC with which the port is associated (known as the `device_id`)

For example, to identify interface `ethernet 1/1` to P4Runtime, you can configure the `interface_id` for the `ethernet 1/1` port, and configure a `device_id` identifying the line card and ASIC associated with the `ethernet 1/1` port. The `device_id,interface_id` tuple uniquely identifies interface `ethernet 1/1`.

The P4Runtime client uses a lookup table consisting of the `device_id,interface_id` tuple → interface-name (as specified by the device) to translate where packets are to be sent to, or populate where a packet was received.

5.1.1 Configuring a port identifier for P4Runtime

Procedure

The `interface_id` part of the `device_id,interface_id` tuple uniquely identifies a port in the SR Linux chassis to a P4Runtime client. You can configure the value for `interface_id`. If you do not configure a value for `interface_id`, the port's `ifIndex` value is used by default.

Example

```
--{ candidate shared default }--[  ]--
# info interface ethernet-1/1 p4rt
    interface ethernet-1/1 {
        p4rt {
            id 2002
        }
    }
```

5.1.2 Configuring a device identifier for P4RT

Procedure

The `device_id` identifies a specific line card and ASIC in the chassis. P4Runtime uses the combination of the `device_id` and `interface_id` to identify the specific interface that a packet was received on (in `PacketIn` messages). Note that for identifying the interface that a packet is to be sent via (in `PacketOut` messages), only the `interface_id` is used.

There is no default `device_id` for a line card / ASIC; you must configure the `device_id` value to be used by P4Runtime.

Example

```
--{ candidate shared default }--[  ]--
# info platform linecard 1 forwarding-complex 0
    platform {
        linecard 1 {
            forwarding-complex 0 {
                p4rt {
                    id 10001
                }
            }
        }
    }
```

5.2 Configuring global P4Runtime server settings

Procedure

You can configure global settings for the P4Runtime server. These settings apply to all network-instances where the P4Runtime server is enabled, and to UNIX sockets if enabled. You can configure the following:

- Whether to administratively enable the P4Runtime server globally
- Limit the number of connection attempts per minute by P4Runtime clients

- Limit the number of P4Runtime RPC connections that can be active at one time
- Idle-timeout in seconds for P4Runtime clients

Example

```
--{ candidate shared default }--[  ]--
# info system p4rt-server
    system {
        p4rt-server {
            timeout 14400
            rate-limit 120
            session-limit 40
        }
    }
```

5.3 Configuring the P4Runtime server for a network-instance

Procedure

You can configure settings for the P4Runtime server that apply to individual network-instances. For a specific network-instance, you can set the P4Runtime server for the following:

- Whether to administratively enable the P4Runtime server for the network-instance
- The port the P4Runtime server listens to for the network-instance. By default, this is TCP port 9559.
- IP addresses the P4Runtime server listens on within the network-instance
- TLS profile to secure communication between P4 Runtime clients and SR Linux for the network-instance
- Whether username/password authentication is used for each P4Runtime RPC request

Example

The following example configures settings for the P4Runtime server for two network-instances:

```
--{ candidate shared default }--[  ]--
# info system p4rt-server network-instance default
    system {
        p4rt-server {
            network-instance blue {
                admin-state enable
                use-authentication true
                tls-profile tls-profile-1
                source-address [
                    192.168.0.1
                ]
            }
            network-instance red {
                admin-state enable
                use-authentication true
                port 9449
                tls-profile tls-profile-2
                source-address [
                    192.168.0.22
                ]
            }
        }
```

```
    }
```

5.4 Configuring the P4Runtime server for UNIX sockets

Procedure

You can configure the following settings for the P4Runtime server that apply to UNIX sockets:

- Whether to administratively enable the P4Runtime server for UNIX sockets
- TLS profile to secure communication between P4 Runtime clients and SR Linux for UNIX sockets
- Whether username/password authentication is used for each P4Runtime RPC request

Example

The following example configures settings for the P4Runtime server for UNIX sockets:

```
--{ candidate shared default }--[  ]--
# info system p4rt-server unix-socket
    system {
        p4rt-server {
            unix-socket {
                admin-state enable
                use-authentication true
                tls-profile tls-profile-3
            }
        }
    }
```

5.5 Disconnecting P4Runtime clients

Procedure

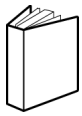
You can use a **tools** command to manually disconnect P4Runtime clients from the server.

To do this, get the identifier for the P4Runtime client from the **info from state system p4rt-server** command, then enter the following command to disconnect the client:

Example

```
-{ running }--[  ]--
# tools system p4rt-server client 4053 disconnect
```

Customer document and product support



Customer documentation

Customer documentation welcome page

Technical support

Product support portal

Documentation feedback

Customer documentation feedback