



# Recommendations

- Huawei Learning Website
  - <http://learning.huawei.com/en>
- Huawei e-Learning
  - <https://ilearningx.huawei.com/portal/#/portal/EBG/51>
- Huawei Certification
  - [http://support.huawei.com/learning/NavigationAction!createNavi?navId=\\_31&lang=en](http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=en)
- Find Training
  - [http://support.huawei.com/learning/NavigationAction!createNavi?navId=\\_trainingsearch&lang=en](http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=en)



## More Information

- Huawei learning APP



Huawei Certification

**HCIA-**

**Routing&Switching**

**INTERMEDIATE**

**Huawei Networking Technology and Device**



**Huawei Technologies Co.,Ltd.**

**Copyright © Huawei Technologies Co., Ltd. 2019.**

**All rights reserved.**

Huawei owns all copyrights, except for references to other parties. No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

**Trademarks and Permissions**



HUAWEI and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

**Notice**

The information in this manual is subject to change without notice. Every effort has been made in the preparation of this manual to ensure accuracy of the contents, but all statements, information, and recommendations in this manual do not constitute the warranty of any kind, express or implied.



**Huawei Certification**

**HCIA-Routing&Switching Huawei Networking Technology**

**and Device**

**Intermediate**

**Version 2.5**

## Huawei Certification System

Relying on its strong technical and professional training and certification system and in accordance with customers of different ICT technology levels, Huawei certification is committed to providing customers with authentic, professional certification, and addresses the need for the development of quality engineers that are capable of supporting Enterprise networks in the face of an ever changing ICT industry. The Huawei certification portfolio for routing and switching (R&S) is comprised of three levels to support and validate the growth and value of customer skills and knowledge in routing and switching technologies.

The Huawei Certified Network Associate (HCIA) certification level validates the skills and knowledge of IP network engineers to implement and support small to medium-sized enterprise networks. The HCIA certification provides a rich foundation of skills and knowledge for the establishment of such enterprise networks, along with the capability to implement services and features within existing enterprise networks, to effectively support true industry operations.

HCIA certification covers fundamentals skills for TCP/IP, routing, switching and related IP network technologies, together with Huawei data communications products, and skills for versatile routing platform (VRP) operation and management.

The Huawei Certified Network Professional (HCIP-R&S) certification is aimed at enterprise network engineers involved in design and maintenance, as well as professionals who wish to develop an in depth knowledge of routing, switching, network efficiency and optimization technologies. HCIP-R&S consists of three units including Implementing Enterprise Routing and Switching Network (IERS), Improving Enterprise Network Performance (IENP), and Implementing Enterprise Network Engineering Project (IEEP), which includes advanced IPv4 routing and switching technology principles, network security, high availability and QoS, as well as application of the covered technologies in Huawei products.

The Huawei Certified Internet Expert (HCIE-R&S) certification is designed to imbue engineers with a variety of IP network technologies and proficiency in maintenance, for the diagnosis and troubleshooting of Huawei products, to equip engineers with in-depth competency in the planning, design and optimization of large-scale IP networks.

# **CONTENTS**

<b>Link Aggregation .....</b>	<b>1</b>
<b>VLAN Principles.....</b>	<b>14</b>
<b>VLAN Routing.....</b>	<b>43</b>
<b>Principle and Configuration of HDLC and PPP .....</b>	<b>57</b>
<b>Principle and Configuration of PPPoE.....</b>	<b>85</b>
<b>Network Address Translation .....</b>	<b>105</b>
<b>Access Control Lists.....</b>	<b>125</b>
<b>AAA.....</b>	<b>139</b>
<b>Securing Data with IPSec VPN.....</b>	<b>151</b>
<b>Generic Routing Encapsulation .....</b>	<b>168</b>
<b>Simple Network Management Protocol .....</b>	<b>184</b>
<b>Introducing IPv6 Networks.....</b>	<b>197</b>
<b>IPv6 Routing Technologies .....</b>	<b>214</b>

<b>IPv6 Application Services-DHCPv6 .....</b>	<b>225</b>
<b>MPLS Basic Principle .....</b>	<b>240</b>
<b>SR Basic Principle.....</b>	<b>266</b>



# Link Aggregation

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- As a means of optimizing the throughput of data, link aggregation enables the binding of multiple physical interfaces into a single logical pipe. This effectively introduces solutions for providing higher utilization of available links, as well as extended resilience in the event that failure of individual links were to occur. Engineers are required to have a clear understanding of the conditions that define the behavior of link aggregation and the skills and knowledge for its application, to ensure effective link aggregation solutions can be applied to enterprise networks.

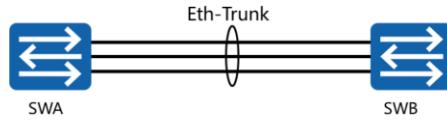


## Objectives

- Upon completion of this section, you will be able to:
  - Explain the use of link aggregation in the enterprise network.
  - Describe the various forms of link aggregation supported.
  - Configure link aggregation solutions.



## Link Aggregation

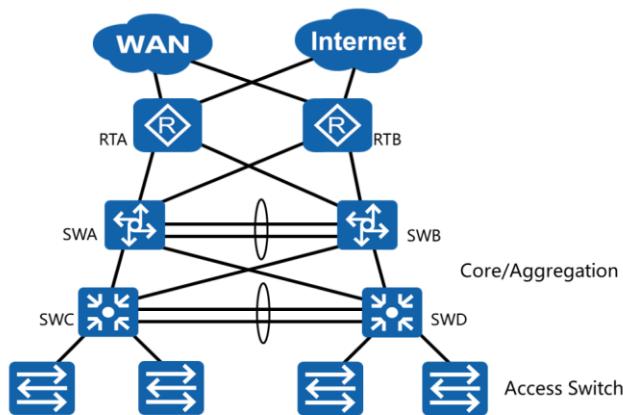


- Link Aggregation provides for increased bandwidth, enhanced reliability and support of load balancing.

- Link aggregation refers to the implementation of a trunk link that acts as a direct point-to-point link, between two devices such as peering routers, switches, or a router and switch combination at each end of the link. The link aggregation comprises of links that are considered members of an Ethernet trunk, and build an association which allows the physical links to operate as a single logical link. The link aggregation feature supports high availability by allowing the physical link of a member interface to switch traffic to another member link in the event that a particular interface fails. In aggregating the links, the bandwidth of a trunk interface is combined, equaling the sum of the bandwidth of all member interfaces, to enable an effective bandwidth increase for traffic over the logical link. Link aggregation can also implement load balancing on a trunk interface. This enables the trunk interface to disperse traffic among its member interfaces, and then transmit the traffic over the member links to the same destination, thus minimizing the likelihood of network congestion.



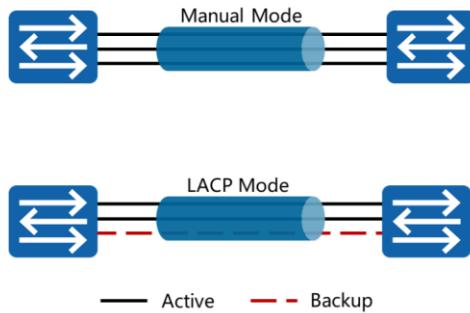
## Application in the Enterprise Network



- Application is made at critical points to enhance throughput.



## Link Aggregation Modes

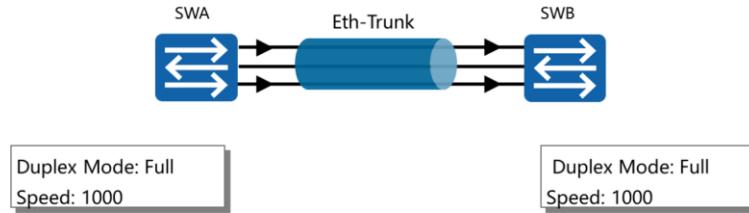


- In manual mode all links load balance and are forwarding.
- LACP mode supports backup links for redundancy.

- Link aggregation supports two modes of implementation, a manual load balancing mode and static LACP mode. In load balancing mode, member interfaces are manually added to a link aggregation group (LAG). All of the interfaces configured with load balancing are set in a forwarding state. The AR2200 can perform load balancing based on destination MAC addresses, source MAC addresses, exclusive-OR of the source and destination MAC addresses, source IP addresses, destination IP addresses, or Exclusive-OR of source and destination IP addresses. The manual load balancing mode does not use the Link Aggregation Control Protocol (LACP), therefore the AR2200 can use this mode if the peer device does not support LACP.
- In static LACP mode, devices at two ends of a link negotiate aggregation parameters by exchanging LACP packets. After the negotiation is complete, the two devices determine the active interface and the inactive interface. In this mode, it is necessary to manually create an Eth-Trunk and add members to it. LACP negotiation determines which interfaces are active and which ones are inactive. The static LACP mode is also referred to as M:N mode, where M signifies the active member links which forward data in a load balancing mode, and N represents those links inactive but providing redundancy. If an active link fails, data forwarding is switched to the backup link with the highest priority, and the status of the backup link changes to active. In static LACP mode, some links may function as backup links, whereas all member interfaces work in a forwarding state in manual load balancing mode, and represents the main difference between the two modes.



## Data Flow Control



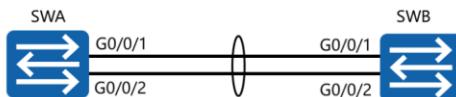
- Data flow sequence must be maintained over member links.
- Consistency of physical member interfaces must be maintained.

- As a logical interface for binding multiple physical interfaces and relaying upper-layer data, a trunk interface must ensure that all parameters of the physical interfaces (member interfaces) on both ends of the trunk link be consistent. This includes the number of physical interfaces, the transmission rates and duplex modes of the physical interfaces, and the traffic-control modes of the physical interfaces, for which it should be noted that member interfaces can be layer 2 or layer 3 interfaces. Where the interface speed is not consistent, it is still possible for the trunk link to operate, however the interfaces operating at a lower rate are likely to experience loss of frames.
- In addition, the sequence of the data flow must be unchanged. A data flow can be considered as a group of frames with the same MAC address and IP address. For example, the telnet or FTP connection between two devices can be considered as a data flow. If the trunk interface is not configured, frames that belong to a data flow can still reach their destination in the correct order because data flows are transmitted over a single physical link. When the trunk technology is used, multiple physical links are bound to the same trunk link, and frames are transmitted along these physical links. If the first frame is transmitted over one physical link, and the second frame is transmitted over another physical link, it is possible that the second frame may reach the destination earlier than the first frame.

- To prevent the disorder of frames, a frame forwarding mechanism is used to ensure that frames in the same data flow reach the destination in the correct sequence. This mechanism differentiates data flows based on their MAC addresses or IP addresses. In this manner, frames belonging to the same data flow are transmitted over the same physical link. After the frame forwarding mechanism is used, frames are transmitted based on the following rules:
- Frames with the same source MAC addresses are transmitted over the same physical link.
- Frames with the same destination MAC addresses are transmitted over the same physical link.
- Frames with the same source IP addresses are transmitted over the same physical link.
- Frames with the same destination IP addresses are transmitted over the same physical link.
- Frames with the same source and destination MAC addresses are transmitted over the same physical link.
- Frames with the same source and destination IP addresses are transmitted over the same physical link.



## L2 Link Aggregation Configuration



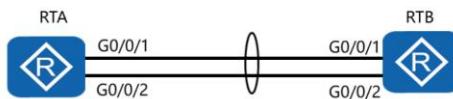
```
[SWA]interface Eth-Trunk 1
[SWA-Eth-Trunk1]interface GigabitEthernet0/0/1
[SWA-GigabitEthernet0/0/1]eth-trunk 1
[SWA-GigabitEthernet0/0/1]interface GigabitEthernet0/0/2
[SWA-GigabitEthernet0/0/2]eth-trunk 1
```

- Link Aggregation requires the binding of the physical member interfaces to the Eth-trunk.

- Establishment of Link Aggregation is achieved using the interface Eth-trunk <trunk-id> command. This command creates an Eth-Trunk interface and allows for the Eth-Trunk interface view to be accessed. The trunk-id is a value used to uniquely identify the Eth-trunk, and can be any integer value from 0 through to 63. If the specified Eth-Trunk already exists, it is possible to directly enter the Eth-Trunk interface view by using the interface Eth-trunk command. An Eth-Trunk can only be deleted if the Eth-Trunk does not contain any member interfaces. When adding an interface to an Eth-Trunk, member interfaces of a layer 2 Eth-Trunk must be layer 2 interfaces, and member interfaces of a layer 3 Eth-Trunk must be layer 3 interfaces. An Eth-Trunk can support a maximum of eight member interfaces. A member interface cannot have any service or static MAC address configured. Interfaces added to an Eth-Trunk should be hybrid interfaces (the default interface type). An Eth-Trunk interface cannot have other Eth-Trunk interfaces as member interfaces. An Ethernet interface can be added to only one Eth-trunk interface.
- To add the Ethernet interface to another Eth-trunk, the Ethernet interface must be deleted from the current Eth-Trunk first. Member interfaces of an Eth-trunk must be the same type, for example, a Fast Ethernet interface and a Gigabit Ethernet interface cannot be added to the same Eth-trunk interface. The peer interface directly connected to a member interface of the local Eth-Trunk must also be added to an Eth-Trunk, otherwise the two ends cannot communicate. When member interfaces have different rates, the interfaces with lower rates may become congested and packet loss may occur. After an interface is added to an Eth-Trunk, MAC address learning is performed by the Eth-Trunk rather than the member interfaces.



## L3 Link Aggregation Configuration



```
[RTA] interface eth-trunk 1
[RTA-Eth-Trunk1]undo portswitch
[RTA-Eth-Trunk1]ip address 100.1.1.1 24
[RTA-Eth-Trunk1]quit
[RTA] interface GigabitEthernet 0/0/1
[RTA-GigabitEthernet0/0/1]eth-trunk 1
[RTA-GigabitEthernet0/0/1] quit
[RTA] interface GigabitEthernet0/0/2
[RTA-GigabitEthernet0/0/2]eth-trunk 1
[RTA-GigabitEthernet0/0/2] quit
```

- In order to configure layer 3 Link Aggregation on an Ethernet trunk link, it is necessary to transition the trunk from layer 2 to layer 3 using the *undo portswitch* command under the Eth-trunk logical interface. Once the *undo portswitch* command has been performed, an IP address can be assigned to the logical interface and the physical member interfaces that are to be associated with the Ethernet trunk link can be added.



## Displaying Aggregation

```
[RTA]display interface eth-trunk 1
Eth-Trunk1 current state : UP
Line protocol current state : UP
.....
-----
PortName          Status      Weight
-----
GigabitEthernet0/0/1    UP         1
GigabitEthernet0/0/2    UP         1
-----
The Number of Ports in Trunk : 2
The Number of UP Ports in Trunk : 2
```

- Two member link ports have been assigned to Eth-trunk 1.

- Using the *display interface eth-trunk <trunk-id>* command it is possible to confirm the successful implementation of Link Aggregation between the two peering devices. The command can also be used to collect traffic statistics and locate faults on the interface.
- The current state of the Eth-trunk is set to UP, signaling that the interface is operating normally. Where the interface shows as down, this signals that an error has occurred at the physical layer, whereas an administratively down error reflects that the *shutdown* command has been used on the interface. The specific error in the event of a failure can be discovered by verifying the status of the ports, for which all ports are expected to show an UP status. Load balancing is supported when the weight of all links is considered equal.



## Summary

- If an administrator attempts to add a Gigabit Ethernet and Fast Ethernet interface to the same Eth-trunk interface, what will occur?
- In order to establish backup member links, which mode of link aggregation should be used?

- A Fast Ethernet interface and a Gigabit Ethernet interface cannot be added to the same Eth-trunk interface, any attempt to establish member links of different types will result in an error specifying that the trunk has added a member of another port-type. It should be noted that the S5700 series switch supports Gigabit Ethernet interfaces only, however this behavior can be applied to other models including the S3700 switch.
- Only the LACP mode is capable of supporting backup member links and therefore should be used if backup links are required.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# VLAN Principles

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- A Virtual Local Area Network (VLAN) represents a form of administrative network that defines a logical grouping of hosts or end system devices that are not limited to a physical location, and may be defined based on a wide range of parameters that allow for a greater flexibility in the way that logical groups are defined. The application of VLAN technology has expanded to support many aspects of enterprise networking as a means of logical data flow management and isolation.

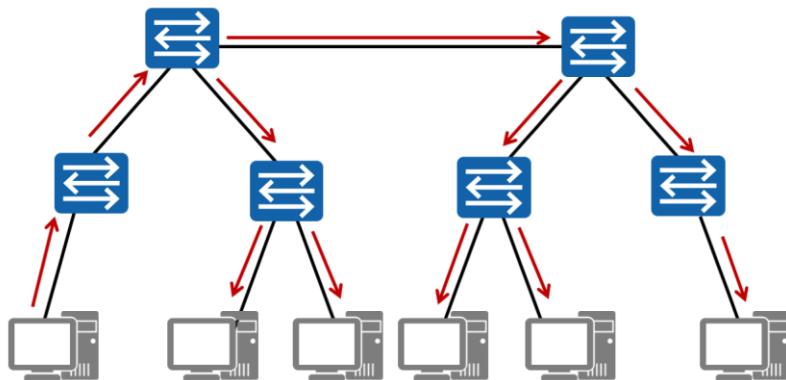


## Objectives

- Upon completion of this section, you will be able to:
  - Explain the application of VLAN tagging.
  - Describe the different port link types and characteristics.
  - Successfully establish port based VLANs.



## LAN Limitations

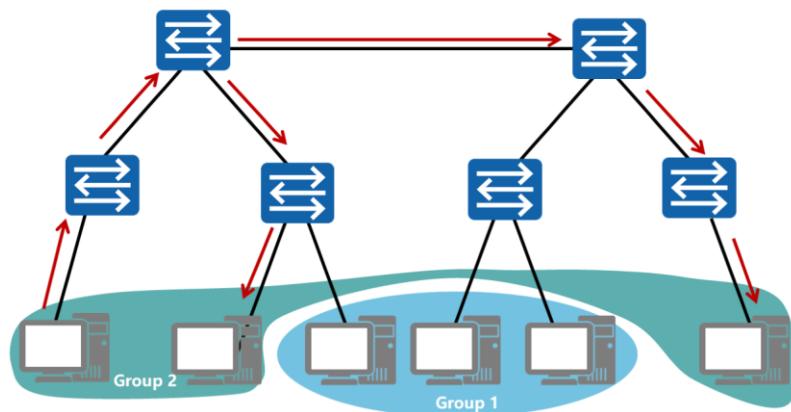


- No broadcast domain to manage expanding local networks.

- As local networks expand, traffic increases and broadcasts become more common. There are no real boundaries within such an expanding network, causing interrupts and growing traffic utilization to occur. Traditionally, the alternative option was to implement a layer three device within the local network to generate broadcast domains, however in doing so additional expense was incurred and the forwarding behavior of such devices did not provide as efficient throughput as found with switches, leading to bottlenecks at transit points between broadcast domains.



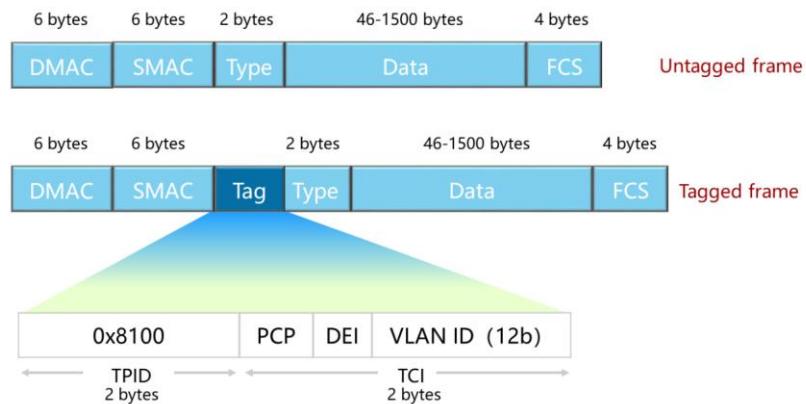
## VLAN Technology



- A VLAN enables logical isolation of traffic at the data link layer.



## VLAN Frame Format

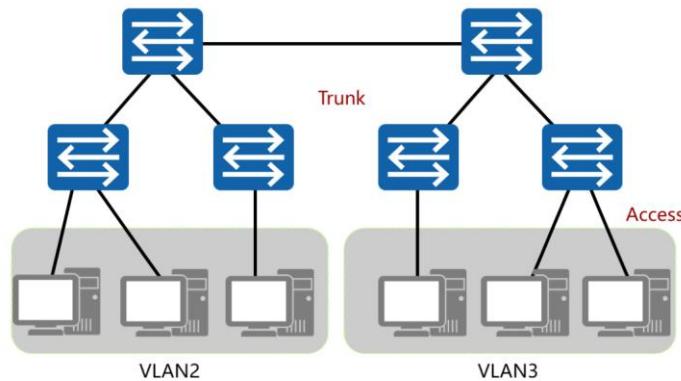


- A VLAN tag is inserted to distinguish frames for each VLAN.

- VLAN frames are identified using a tag header which is inserted into the Ethernet frame as a means of distinguishing a frame associated with one VLAN from frames of another. The VLAN tag format contains a Tag Protocol Identifier (TPID) and associated Tag Control Information (TCI). The TPID is used to identify the frame as a tagged frame, which currently only refers to the IEEE 802.1Q tag format, for which a value of 0x8100 is used to identify this format. The TCI contains fields that are associated with the tag format type.
- The Priority Code Point (PCP) is a form of traffic classification field that is used to differentiate one form of traffic from another so as to prioritize traffic generally based on a classification such as voice, video, data etc. This is represented by a three bit value allowing a range from 0-7, and can be understood based on general 802.1p class of service (CoS) principles. The Drop Eligibility Indicator (DEI) represents a single bit value that exists in either a True or False state to determine the eligibility of a frame for discarding in the event of congestion.
- The VLAN ID indicates the VLAN with which the frame is associated, represented as a 12 bit value. VLAN ID values range from 0x000 through to 0xFFFF and for which the two upper and lower values are reserved, allowing 4094 possible VLAN Combinations. Huawei VRP implementation of VLANs uses VLAN 1 as the default VLAN (PVID) as based on IEEE802.1Q standards.



## Link Types

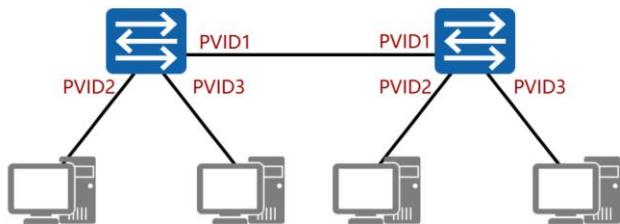


- A trunk represents a backbone for the transmission of VLAN traffic between switches.

- VLAN links can be classified into two types, an access link type and a trunk link type. The access link refers to the link between an end system and a switch device participating in VLAN tagging, the link between host terminals and switches are all access links. A trunk link refers to the link over which VLAN tagged frames are likely to be carried. The links between switches are generally understood to be trunk links.



## Port VLAN ID

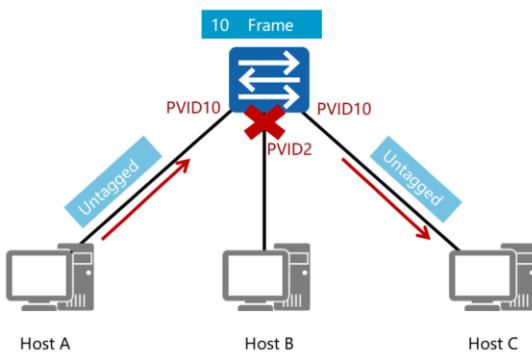


- PVID represents the default VLAN for each interface.
- The PVID is set to VLAN 1 for all ports by default.

- Each interface of a device participating in VLAN tagging will be associated with a VLAN. The default VLAN for the interface is recognized as the Port VLAN ID (PVID). This value determines the behavior that is applied to any frames being received or transmitted over the interface.



## Port Types – Access

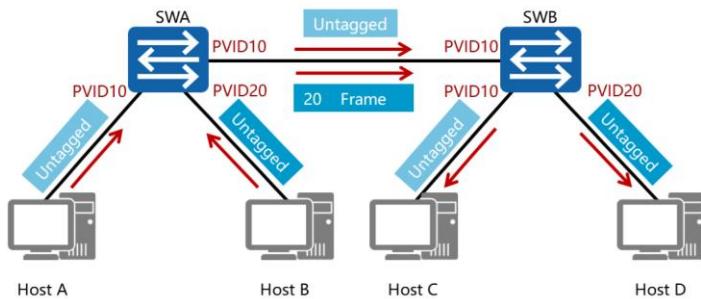


- Access ports remove VLAN tags before forwarding frames.

- Access ports associate with access links, and frames that are received will be assigned a VLAN tag that is equal to the Port VLAN ID (PVID) of the interface. Frames being transmitted from an interface will typically remove the VLAN tag before forwarding to an end system that is not VLAN aware. If the tag and the PVID vary however, the frame will not be forwarded and therefore discarded. In the example a frame (untagged) is forwarded to the interface of the switch, which can be understood to forward to all other destinations.
- Upon receiving the frame, the switch will associate the frame with VLAN 10 based on the PVID of the interface. The switch is able to identify at the port interface the PVID and make a decision as to whether the frame can be forwarded. In the case of Host C the PVID matches the VLAN ID in the VLAN tag, for which the tag is removed and the frame forwarded. For Host B however the frame and the PVID differ, and therefore the frame is restricted from being forwarded to this destination.



## Port Types – Trunk

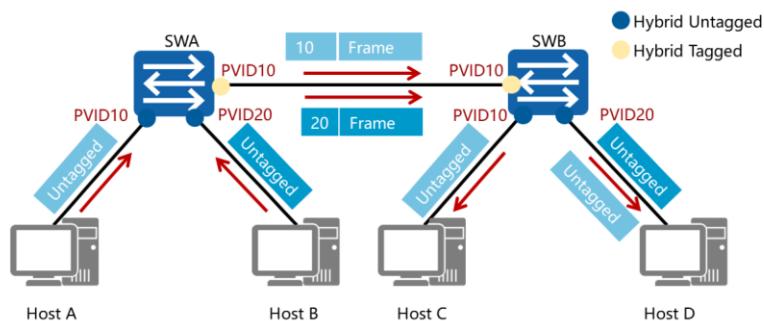


- Frames carried over a trunk link may be tagged or untagged.
- All VLANs must be permitted before being carried over a trunk.

- For trunk ports that are associated with trunk links, the Port VLAN ID (PVID) will identify which VLAN frames are required to carry a VLAN tag before forwarding, and which are not. The example demonstrates a trunk interface assigned with a PVID of 10, for which it should be assumed that all VLANs are permitted to traverse the trunk link. Only frames associated with VLAN 10 will be forwarded without the VLAN tag, based on the PVID. For all other VLAN frames, a VLAN tag must be included with the frame and be permitted by the port before the frame can be transmitted over the trunk link. Frames associated with VLAN 20 are carried as tagged frames over the trunk link.



## Port Types – Hybrid

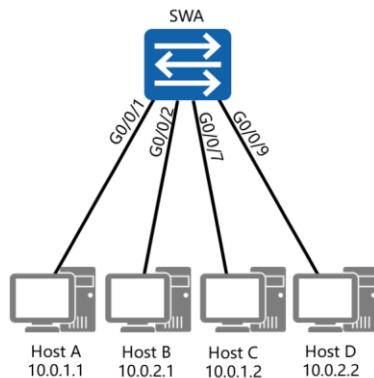


- Hybrid ports are defined as either tagged or untagged.
- VLAN communication can be managed on a port by port basis.

- Hybrid represents the default port type for Huawei devices supporting VLAN operation and provides a means of managing the tag switching process associated for all interfaces. Each port can be considered as either a tagged port or an untagged port. Ports which operate as access ports (untagged) and ports which operate as trunk ports (tagged).
- Ports which are considered untagged will generally receive untagged frames from end systems, and be responsible for adding a tag to the frame based on the Port VLAN ID (PVID) of the port. One of the key differences is in the hybrid port's ability to selectively perform the removal of VLAN tags from frames that differ from the PVID of the port interface. In the example, Host D is connected to a port which specifies a Port VLAN ID of 20, whilst at the same time is configured to allow for the removal of the tag from frames received from VLAN 10, thereby allowing Host D to receive traffic from both VLANs 10 & 20.
- Hybrid Ports that are tagged will operate in a similar manner as a regular trunk interface, however one major difference exists. VLAN frames that both match the PVID and are permitted by the port will continue be tagged when forwarded.



## VLAN Assignment Methods



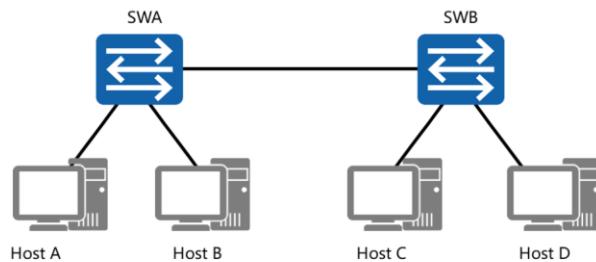
Assignment Method	VLAN 5	VLAN 10
Port based	G0/0/1, G0/0/7	G0/0/2 G0/0/9
MAC based	00-01-02-03-04-AA 00-01-02-03-04-CC	00-01-02-03-04-BB 00-01-02-03-04-DD
IP Subnet based	10.0.1.*	10.0.2.*
Protocol based	IP	IPX
Policy based	10.0.1.* + G0/0/1+ 00-01-02-03-04-AA	10.0.2.* + G0/0/2 + 00-01-02-03-04-BB

- Five methods of VLAN assignment are possible.
- Port based VLAN assignment is the default assignment method.

- VLAN assignment can be implemented based on one of five different methods, including Port based, MAC based, IP Subnet based, Protocol based and Policy based implementations. The port based method represents the default and most common method for VLAN assignment. Using this method, VLANs are classified based on the port numbers on a switching device. The network administrator configures a Port VLAN ID (PVID), representing the default VLAN ID for each port on the switching device. When a data frame reaches a port, it is marked with the PVID if the data frame carries no VLAN tag and the port is configured with a PVID. If the data frame carries a VLAN tag, the switching device will not add a VLAN tag to the data frame even if the port is configured with a PVID.
- Using the MAC address assignment method, VLANs are classified based on the MAC addresses of network interface cards (NICs). The network administrator configures the mappings between MAC addresses and VLAN IDs. In this case, when a switching device receives an untagged frame, it searches the MAC-VLAN table for a VLAN tag to be added to the frame according to the MAC address of the frame. For IP subnet based assignment, upon receiving an untagged frame, the switching Device adds a VLAN tag to the frame based on the IP address of the packet header.
- Where VLAN classification is based on protocol, VLAN IDs are allocated to packets received on an interface according to the protocol (suite) type and encapsulation format of the packets. The network administrator configures the mappings between types of protocols and VLAN IDs. The Policy based assignment implements a combination of criteria for assignment of the VLAN tag, including the IP subnet, port and MAC address, in which all criteria must match before the VLAN is assigned.



## Creating VLANs



```
[SWA]vlan 10
[SWA-vlan10]quit
[SWA]vlan batch 2 to 3
Info: This operation may take a few seconds. Please wait for a
moment...done.
```

- The implementation of VLANs begins with the creation of the VLAN on the switch. The `vlan<vlan-id>` command is used to initially create the the VLAN on the switch which can be understood to exist once the user enters the VLAN view for the given vlan as demonstrated in the configuration example. The VLAN ID ranges from 1 to 4094 and where it is necessary to create multiple VLANs for a switch, the `vlan batch <vlan-id1 to vlan-id2>` command can be used where contiguous VLAN ranges need to be created and `vlan batch &<1-4094>` command used where “&” represents a space between non-contiguous VLAN ranges. All ports are associated with VLAN 1 as the default VLAN by default, and therefore forwarding is unrestricted.



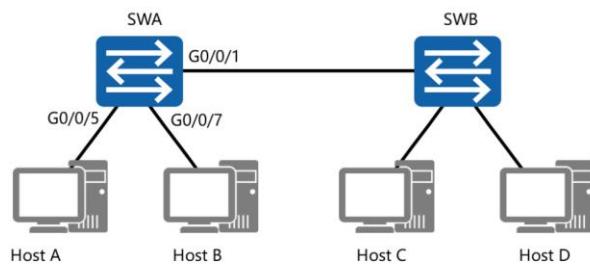
## Creating VLANs

```
[SWA]display vlan
The total number of vlans is : 4
-----
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
-----
VID Type Ports
-----
1 common UT:GE0/0/1(U) .....
2 common
3 common
10 common
.....
```

- Once the VLANs have been created, the creation can be verified using the `display vlan` command. The command allows information about all VLANs to be specified, and if no parameter is specified, brief information about all VLANs is displayed. Additional parameters include `display vlan <vlan-id> verbose` command, used to display detailed information about a specified VLAN, including the ID, type, description, and status of the VLAN, status of the traffic statistics function, interfaces in the VLAN, and mode in which the interfaces are added to the VLAN. The `display vlan <vlan-id> statistics` command, allows for the view of traffic statistics on interfaces for a specified VLAN. The `display vlan summary` command, provides a summary of all VLANs in the system.



## Setting the Port Link Type

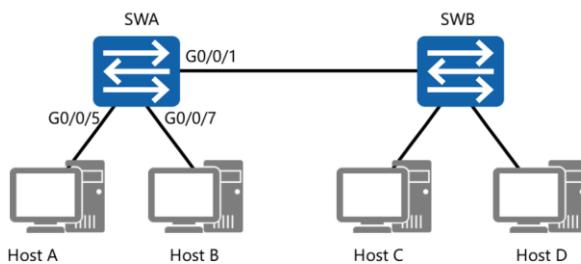


```
[SWA] interface GigabitEthernet 0/0/1
[SWA-GigabitEthernet0/0/1]port link-type trunk
[SWA-GigabitEthernet0/0/1]quit
[SWA] interface GigabitEthernet 0/0/5
[SWA-GigabitEthernet0/0/5]port link-type access
```

- The configuration of the port link type is performed in the interface view for each interface on a VLAN active switch. The default port link type on Huawei switch devices is hybrid. The *port link-type <type>* command is used to configure the port link type of the interface where the type can be set as access, trunk or hybrid. A fourth QinQ option exists but is considered outside of the scope of this course. It should also be noted that in the displayed configuration if no port type is displayed, the default hybrid port link type is configured. Prior to changing the interface type, it is also necessary to restore the default VLAN configuration of the interface so that the interface belongs to only the default VLAN 1.



## Assigning Ports to VLANs



```
[SWA]vlan 2  
[SWA-vlan2]port GigabitEthernet 0/0/7  
[SWA-vlan2]quit  
[SWA]interface GigabitEthernet 0/0/5  
[SWA-GigabitEthernet0/0/5]port link-type access  
[SWA-GigabitEthernet0/0/5]port default vlan 3
```



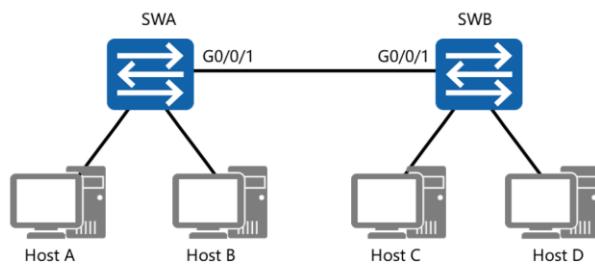
## Verifying VLAN Assignment

```
[SWA]display vlan
The total number of vlans is : 4
-----
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
-----
VID Type Ports
-----
1 common UT:GE0/0/1(U) ....
2 common UT:GE0/0/7(D)
3 common UT:GE0/0/5 (U)
10 common
.....
```

- The `display vlan` command can be used to verify the changes made to the configuration and confirm the association of port interfaces with the VLANs to which the ports have been assigned. In the display example port interfaces Gigabit Ethernet 0/0/5 and Gigabit Ethernet 0/0/7 can be identified as being associated with VLANs 2 and 3 respectively. The UT value identifies that the port is considered untagged either through assigning of the port link type as an access port or as an untagged hybrid port. The current state of the link can also be determined as either up (U) or down (D).



## Forwarding Over the Trunk



```
[SWA-GigabitEthernet0/0/1]port link-type trunk  
[SWA-GigabitEthernet0/0/1]port trunk pvid vlan 10  
[SWA-GigabitEthernet0/0/1]port trunk allow-pass vlan 2 3
```

- The assigning of the port link type of trunk interfaces enables the trunk to support the forwarding of VLAN frames for multiple VLANs between switches, however in order for frames to be carried over the trunk interface, permissions must be applied. The port trunk allow-pass vlan <vlan-id> command is used to set the permission for each VLAN, where vlan-id refers to the VLANs to be permitted. It is also necessary that the PVID for the trunk interface be included in the command to enable untagged traffic to be carried over the trunk link. The example demonstrates the changing of the default Port VLAN ID (PVID) for the interface to 10 and the applying of permission for VLANs 2 and 3 over the trunk link. In this case, any frames associated with VLAN 10 will not be carried over the trunk even though VLAN 10 is now the default VLAN for the trunk port. The command port trunk allow-pass vlan all can be used to allow all VLANs to traverse the trunk link.



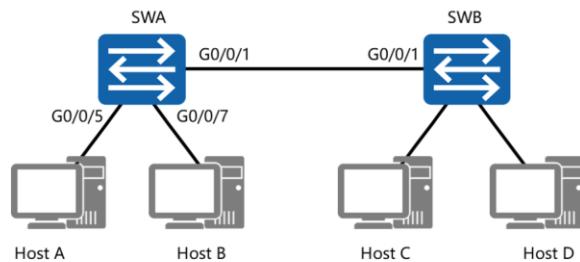
## Forwarding Over the Trunk

```
[SWA]display vlan
The total number of vlans is : 4
-----
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
-----
VID Type Ports
-----
1 common UT:GE0/0/1(U) ....
2 common UT:GE0/0/7(D) TG:GE0/0/1(U)
3 common UT:GE0/0/5(U) TG:GE0/0/1(U)
10 common
.....
```

- The changes to the VLAN permissions can again be monitored through the `display vlan` command, for which the application of VLANs over the trunk link are reflected. The TG value identifies that VLANs have been associated with a tagged interface either over a trunk or tagged hybrid port interface. In the display example, VLANs 2 and 3 have been given permission to traverse the tagged interface Gigabit Ethernet 0/0/1, an interface that is currently active.



## Configuring Hybrid Ports

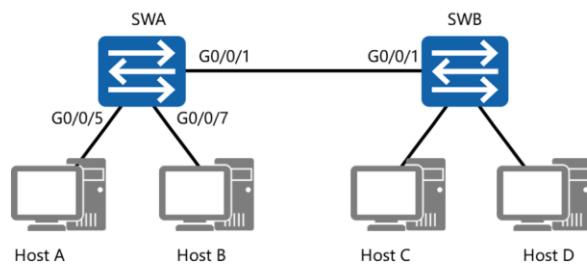


```
[SWA-GigabitEthernet0/0/5]port link-type hybrid  
[SWA-GigabitEthernet0/0/5]port hybrid pvid vlan 3  
[SWA-GigabitEthernet0/0/5]port hybrid untagged vlan 3  
  
[SWA-GigabitEthernet0/0/7]port link-type hybrid  
[SWA-GigabitEthernet0/0/7]port hybrid pvid vlan 2  
[SWA-GigabitEthernet0/0/7]port hybrid untagged vlan 2
```

- Hybrid port configuration represents the default port type on switch port interfaces and therefore the command port link-type hybrid is generally only necessary when converting the port link type from an access or a trunk port link type. Each port however may require to be associated with a default Port VLAN ID (PVID) over which frames are required to be either tagged or untagged. The port hybrid pvid vlan <vlan-id> command enables the default PVID to be assigned on a port by port basis following which it is also necessary to associate the forwarding behavior for a given port.
- For ports that are to operate as access ports, this is achieved using the port hybrid untagged vlan<vlan-id> command. It should be clearly noted that the use of this command multiple times under the same interface view shall result in the interface being associated with all VLANs specified, with the associated VLAN frames being untagged before forwarding. The undo port hybrid vlan command can be used restore the default VLAN setting of VLAN1 and return to the default untagged mode.



## Configuring Hybrid Ports



```
[SWA-GigabitEthernet0/0/1]port link-type hybrid  
[SWA-GigabitEthernet0/0/1]port hybrid tagged vlan 2 to 3
```

- Trunk links using the hybrid port link-type must enable tagging of VLAN frames before forwarding.

- For ports that are to operate as trunk ports, the *port hybrid tagged vlan <vlan-id>* command is used. It should be clearly noted that the use of this command multiple times under the same interface view shall result in the interface being associated with all VLANs specified, with the associated VLAN frames being tagged before forwarding. In the example the hybrid port interface Gigabit Ethernet 0/0/1 is expected to tag all frames that are associated with VLANs 2 and 3 before such frames are forwarded over the interface.



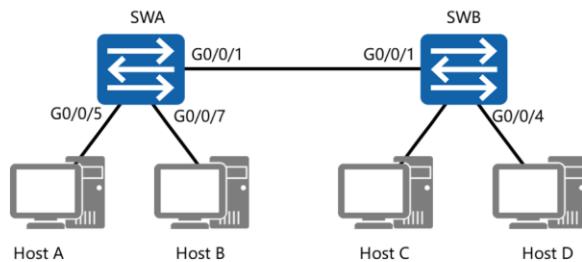
## Configuration Validation

```
[SWA]display vlan
The total number of vlans is : 4
-----
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping; ST:Vlan-
stacking; #: ProtocolTransparent-vlan; *:Management-vlan;
-----
VID Type Ports
-----
1 common UT:GE0/0/1(U) ....
2 common UT:GE0/0/7(D)
          TG:GE0/0/1(U)
3 common UT:GE0/0/5(U)
          TG:GE0/0/1(U)
10 common
....
```

- Through the *display vlan* command, the results of the tagged and untagged hybrid port configuration can be verified. Interface Gigabit Ethernet 0/0/7 has been established as a VLAN 2 untagged interface, while interface Gigabit Ethernet 0/0/5 has been established as an untagged interface associated with VLAN 3. In terms of both VLAN 2 and VLAN 3, frames associated with either VLAN will be carried as a tagged frame over interface Gigabit Ethernet 0/0/1.



## Configuring Hybrid Ports



```
[SWB-GigabitEthernet0/0/4]port link-type hybrid  
[SWB-GigabitEthernet0/0/4]port hybrid pvid vlan 3  
[SWB-GigabitEthernet0/0/4]port hybrid untagged vlan 2 to 3
```

- Hybrid ports can be configured to receive VLAN traffic from multiple VLANs by simply removing the tag at the port interface.

- Switch port interfaces can use the *port hybrid untagged vlan <vlan-id> [to <vlan-id>]* command to apply the untagged behavior on a port interface for multiple VLANs in a single batch command. This behavior enables hybrid interfaces to permit the untagged forwarding of traffic from multiple VLANs to a given end system. All traffic forwarded from the end system is associated with the PVID assigned to the port and tagged respectively.

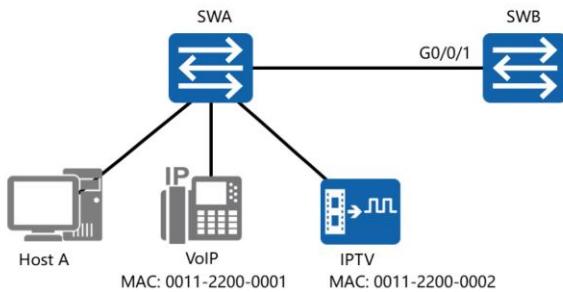


## Configuration Validation

```
[SWB]display vlan
The total number of vlans is : 3
-----
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping; ST:Vlan-
stacking; #: ProtocolTransparent-vlan; *:Management-vlan;
-----
VID Type Ports
-----
1 common UT:GE0/0/1(U) ....
2 common UT:GE0/0/4 (U)
3 common UT:GE0/0/4 (U)
4 ....
```



## Voice VLAN Application

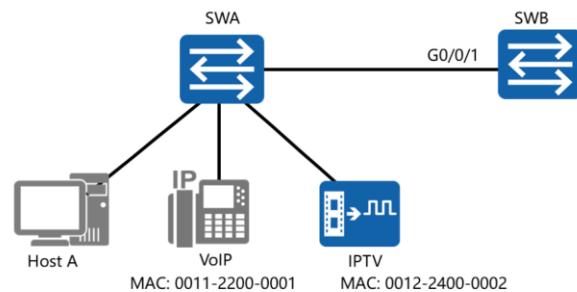


- Voice VLANs are used to distinguish, isolate and prioritize voice traffic over service traffic as a means of quality assurance.

- The growth of IP convergence has seen the integration of multiple technologies that allows High Speed Internet (HSI) services, Voice over IP (VoIP) services, and Internet Protocol Television (IPTV) services to be transmitted over a common Ethernet & TCP/IP network. These technologies originate from networks consisting of different forms of behavior. VoIP originates from circuit switched network technologies that involve the establishment of a fixed circuit between the source and destination, over which a dedicated path is created, ensuring that voice signals arrive with little delay and in a first-in-first-out signal order.
- High Speed Internet operates in a packet switched network involving contention, and packet forwarding with no guarantee of orderly delivery for which packet resequencing is often necessary. Guaranteeing that technologies originating from a circuit switched network concept are capable of functioning over packet switched networks has brought about new challenges. This challenge focuses on ensuring that the services are capable of differentiating voice data from other data. The solution involves VoIP traffic being isolated through different VLANs and being assigned a higher priority to ensure voice quality throughput. Special voice VLANs can be configured on the switch, which allows the switch to assign a pre-configured VLAN ID and a higher priority to VoIP traffic.



## Voice VLAN Configuration



```
[SWB]vlan 2
[SWB-vlan2]interface GigabitEthernet 0/0/1
[SWB-GigabitEthernet0/0/1]voice-vlan 2 enable
[SWB-GigabitEthernet0/0/1]voice-vlan mode auto
[SWB-GigabitEthernet0/0/1]quit
[SWB]voice-vlan mac-address 0011-2200-0000 mask ffff-ff00-0000
```

- Configuration of the voice VLAN involves the configuring of a specified VLAN using the *voice-vlan <vlan-id> enable* command. The voice VLAN can be associated with any VLAN between 2 and 4094. The *voice-vlan mode <mode>* command specifies the working mode, by which a port interface is added to a voice VLAN. This is set by default to occur automatically however can be also achieved manually. The *voice-vlan mac-address <mac-address> mask <mask>* command allows voice packets originating from an IP phone to be identified and associated with the voice VLAN, based on the Organizationally Unique Identifier (OUI), to ultimately allow a higher priority to be given to voice traffic.



## Configuration Validation

```
[SWB]display voice-vlan status
Voice VLAN Configurations:
-----
Voice VLAN ID      : 2
Voice VLAN status   : Enable
Voice VLAN aging time : 1440 (minutes)
Voice VLAN 8021p remark : 6
Voice VLAN dscp remark : 46
-----
Port Information:
-----
Port          Add-Mode Security-Mode Legacy
-----
GigabitEthernet0/0/1    Auto     Security   Disable
```

- The `display voice-vlan status` command allows voice VLAN information to be viewed, including the status, security mode, aging time, and the interface on which the voice VLAN function is enabled. The status determines whether the voice VLAN is currently enabled or disabled. The security-mode can exist in one of two modes, either normal or security. The normal mode allows the interface enabled with voice VLAN to transmit both voice data and service data, but remains vulnerable to attacks by invalid packets. It is generally used when multiple services (HSI, VoIP, and IPTV) are transmitted to a Layer 2 network through one interface, and the interface transmits both voice data and service data. The security mode applied on an interface enabled with voice VLAN checks whether the source MAC address of each packet that enters the voice VLAN matches the OUI. It is applied where the voice VLAN interface transmits ONLY voice data. The security mode can protect the voice VLAN against the attacks by invalid packets, however checking packets occupies certain system resources.
- The Legacy option determines whether the interface can communicate with voice devices of other vendors, where an enabled interface permits this communication. The Add-Mode determines the working mode of the voice VLAN. In auto voice VLAN mode, an interface can be automatically added to the voice VLAN after the voice VLAN function is enabled on the interface, and adds the interface connected to a voice device to the voice VLAN if the source MAC address of packets sent from the voice device matches the OUI. The interface is automatically deleted if the interface does not receive any voice data packets from the voice device within the aging time. In manual voice VLAN mode, an interface must be added to the voice VLAN manually after the voice VLAN function is enabled on the interface.



## Summary

- If a trunk link has a PVID of 5 and the command port trunk allow-pass vlan 2 3 is used, which VLAN traffic will be carried over the trunk?
- What action will be taken by an access port with a PVID of 2 when receiving an untagged frame?

- The PVID on a trunk link defines only the tagging behavior that will be applied at the trunk interface. If the port trunk allow-pass vlan 2 3 command is used, only frames associated with VLAN 2 and VLAN 3 will be forwarded over the trunk link.
- An access port configured with a PVID of 2 will tag all received untagged frames with a VLAN 2 tag. This will be used by the switch to determine whether a frame can be forwarded via other access interfaces or carried over a trunk link.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## VLAN Routing

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The implementation of VLAN technology within an enterprise network effectively establishes broadcast domains that control the scope of traffic. One of the limitations of broadcast domains is that communication at the link layer is hindered between hosts that are not part of the same VLAN. Traditional link layer switches supporting VLANs are not capable of forwarding traffic between these broadcast domains, and therefore routing must be introduced to facilitate communication. The application of VLAN routing when using link layer switches, together with a device capable of routing VLAN traffic is introduced, along with details of how layer three switches capable of network layer operations can enable communication over VLAN defined broadcast domains.

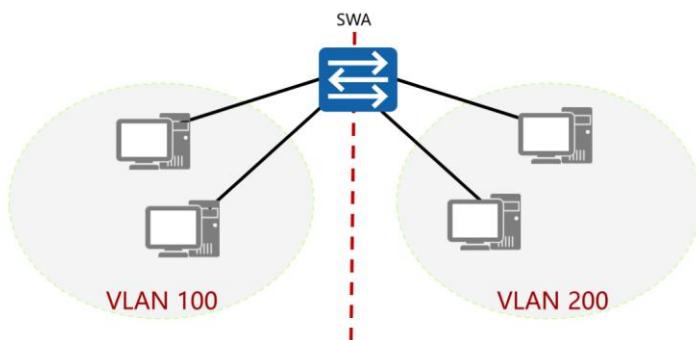


## Objectives

- Upon completion of this section, you will be able to:
  - Explain the purpose of VLAN routing.
  - Explain how VLAN routing is achieved for layer 2 & layer 3 switches.
  - Configure VLAN routing.



## VLAN Disadvantages

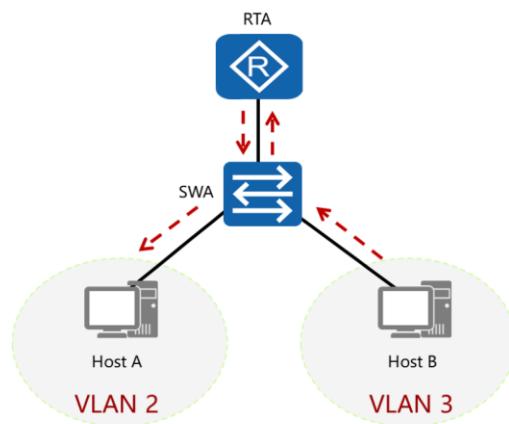


- Attempts to limit broadcast domain size through VLAN implementation isolates users.

- The general principle of VLAN implementation is to isolate networks as a means of minimizing the size of the existing broadcast domain, however in doing so, many users are cut off from other users within other VLAN domains and require that layer three (IP) communication be established in order for those broadcast domains to re-establish communication through reachable routes. The implementation of a layer three switch offers an ideal means for supporting VLAN routing whilst reducing operating costs. One of the constraints however of VLAN routing is the need for strict IP address management.
- Generally however the VLAN routing principle is applicable to small scale networks on which users belong to different network segments and IP addresses of users are seldom changed.



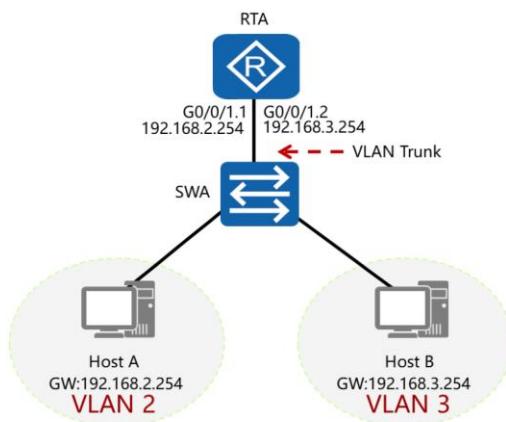
## VLAN Routing



- VLAN frames are routed over a trunk link for port conservation.



## VLAN Routing Features

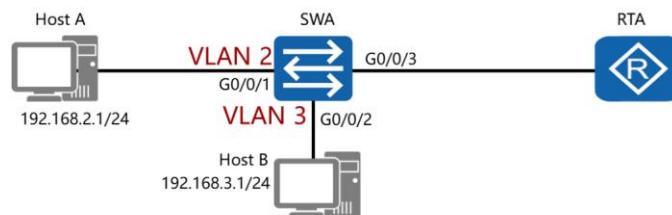


- A single trunk supports VLAN routes by using sub-interfaces.

- In order to allow communication over a single trunk interface, it is necessary to logically segment the physical link using sub-interfaces. Each sub-interface represents a logical link for the forwarding of VLAN traffic before being routed by the router via other logical sub-interfaces to other VLAN destinations. Each sub-interface must be assigned an IP address in the same network segment as the VLAN that it is created for as well as 802.1Q encapsulation to allow for VLAN association as traffic is routed between VLANs.
- It is also necessary to configure the type of the Ethernet port of the switch that connects to the router as either a Trunk or Hybrid link type, and allow frames of the associated VLANs (VLAN 2 & VLAN 3 in this case) to pass.



## VLAN Routing Configuration

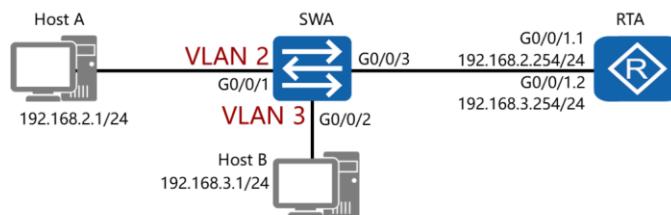


```
[SWA]vlan batch 2 3  
[SWA-GigabitEthernet0/0/1]port link-type access  
[SWA-GigabitEthernet0/0/1]port default vlan 2  
[SWA-GigabitEthernet0/0/2]port link-type access  
[SWA-GigabitEthernet0/0/2]port default vlan 3  
[SWA-GigabitEthernet0/0/3]port link-type trunk  
[SWA-GigabitEthernet0/0/3]port trunk allow-pass vlan all
```

- The trunk link between the switch and the router must be established for support of traffic for multiple VLANs, through the port link-type trunk or port link-type hybrid command as well as the port trunk allow-pass vlan 2 3 or port hybrid vlan 2 3 command respectively. Once the trunk is established, the VLAN sub-interfaces must be implemented to allow the logical forwarding of traffic between VLANs over the trunk link.



## VLAN Routing Configuration



```
[RTA] interface GigabitEthernet0/0/1.1  
[RTA-GigabitEthernet0/0/1.1]dot1q termination vid 2  
[RTA-GigabitEthernet0/0/1.1]ip address 192.168.2.254 24  
[RTA-GigabitEthernet0/0/1.1]arp broadcast enable  
[RTA] interface GigabitEthernet0/0/1.2  
[RTA-GigabitEthernet0/0/1.2]dot1q termination vid 3  
[RTA-GigabitEthernet0/0/1.2]ip address 192.168.3.254 24  
[RTA-GigabitEthernet0/0/1.2]arp broadcast enable
```

- The sub-interface on a router is defined in the interface view using the interface <interface-type interface-number.sub-interface number> command where the sub-interface number represents the logical interface channel within the physical interface. The command dot1q termination vid <vlan-id> is used to perform two specific functions. Where a port receives a VLAN packet, it will initially remove the VLAN tag from the frame and forward this packet via layer three routing.
- For packets being sent out, the port adds a tag to the frame before sending it out, in accordance with the respective VLAN and IP settings for the router's logical interface. Finally the arp-broadcast enable command is applied to each logical interface. This is necessary as the capability for ARP to broadcast on sub-interfaces is not enabled by default. If ARP broadcasts remain disabled on the sub-interface, the router will directly discard packets. The route to the sub-interface generally is considered as a blackhole route in these cases since the packet is effectively lost without a trace. If ARP broadcasts are enabled on the sub-interface, the system is able to construct a tagged ARP broadcast packet and send the packet from the sub-interface.



## VLAN Routing Configuration

```
HostA>ping 192.168.3.1

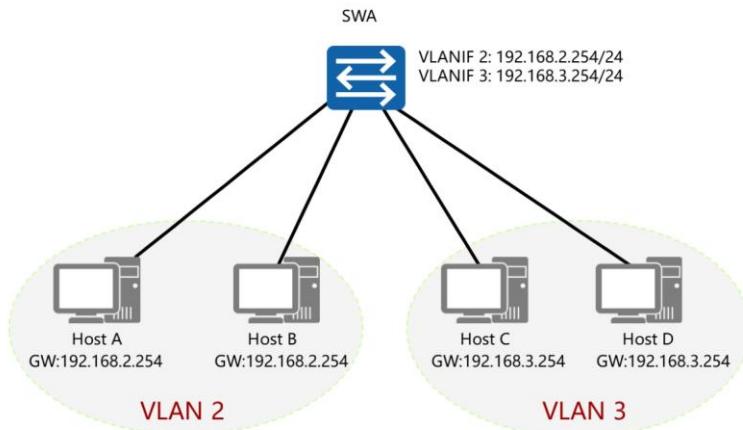
Ping 192.168.3.1: 32 data bytes, Press Ctrl_C to break
From 192.168.3.1: bytes=32 seq=1 ttl=127 time=15 ms
From 192.168.3.1: bytes=32 seq=2 ttl=127 time=15 ms
From 192.168.3.1: bytes=32 seq=3 ttl=127 time=32 ms
From 192.168.3.1: bytes=32 seq=4 ttl=127 time=16 ms
From 192.168.3.1: bytes=32 seq=5 ttl=127 time=31 ms

--- 192.168.3.1 ping statistics ---
5 packet(s) transmitted
5 packet(s) received
0.00% packet loss
round-trip min/avg/max = 15/21/32 ms
```

- Following the configuration of VLAN routing between VLAN 2 and VLAN 3, the ping application can be used to verify reachability. The example demonstrates how Host A (192.168.2.2) in VLAN 2 is capable of reaching Host B (192.168.3.2) in VLAN 3. The TTL reflects that the packet has traversed the router to reach the destination in VLAN 2.



## L3 Switch based VLAN Routing

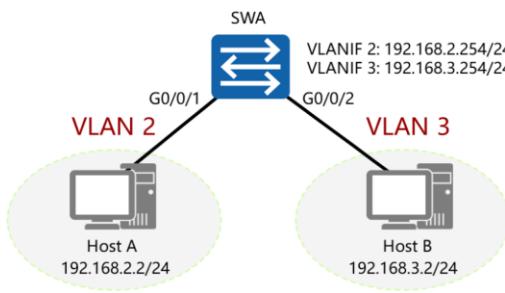


- VLANIF are used by each VLAN as a route gateway.

- The implementation of L3 switches brings about benefits to the process of VLAN routing that are not possible through the use of a router. One of those features is the ability to forward VLAN traffic with very little delay due to support of what is known as line speed forwarding as a result of bottom layer ASIC chips that allow traffic to be forwarded based on hardware rather than software. Along with this is the fact that a single device is used with no trunk link that may otherwise face congestion under heavy traffic loads. VLAN routing when using a layer 3 switch relies on the implementation of VLAN interfaces (VLANIF). If multiple users on a network belong to different VLANs, each VLAN requires a VLANIF that acts as the VLAN gateway and so must associate with an IP address relevant to the network of the VLAN. If a large number of VLANs exist however, this can tally up to a large number of IP addresses being required to support each VLANIF, as well as the hosts that are part of the VLAN with which the VLANIF is associated. Through the VLANIF, routing between different VLANs can be supported.



## L3 Switch Configuration

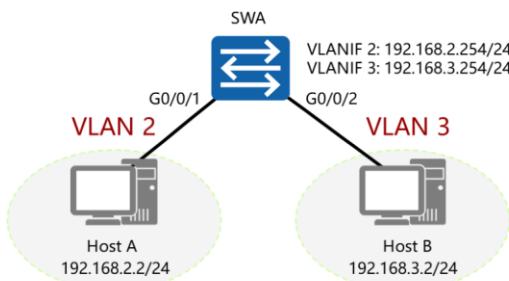


```
[SWA]vlan batch 2 3
[SWA]interface GigabitEthernet 0/0/1
[SWA-GigabitEthernet0/0/1]port link-type access
[SWA-GigabitEthernet0/0/1]port default vlan 2
[SWA]interface GigabitEthernet 0/0/2
[SWA-GigabitEthernet0/0/2]port link-type access
[SWA-GigabitEthernet0/0/2]port default vlan 3
```

- Configuration of VLAN routing on a switch operating at layer 3 requires that the VLANs be initially created and the interfaces be assigned to those respective VLANs. The configuration follows the principles for configuration of VLANs covered as part of the VLAN principles. This involves defining the port link-type for each port and the PVID that is associated with each port interface.



## L3 Switch Configuration



```
[SWA]interface vlanif 2  
[SWA-Vlanif2]ip address 192.168.2.254 24  
[SWA-Vlanif2]quit  
[SWA]interface vlanif 3  
[SWA-Vlanif3]ip address 192.168.3.254 24  
[SWA-Vlanif3]quit
```

- Configuration of VLAN routing is implemented by creating VLAN interfaces that are to operate as gateway interfaces for each VLAN within the layer 3 switch. Entering the VLANIF view is achieved via the *interface vlanif <vlan-id>* command, where the *vlan-id* refers to the associated VLAN. The IP address for the interface should be in the same network segment as the hosts. This IP address shall represent the gateway for the hosts and support the inter-VLAN communication.



## Summary

- What is the purpose of the dot1q termination vid <vlan-id> command?
- What is required to be configured on the switch to allow VLAN traffic to be forwarded to the configured sub-interfaces?

- The dot1q termination vid <vlan-id> command is used to perform two specific functions. Where a port receives a VLAN packet, it will initially remove the VLAN tag from the frame and forward this packet via layer 3 routing. For packets being sent out, the port adds a tag to the packet before sending it out, in accordance with the respective VLAN and IP settings for the routers logical interface.
- The switch must be configured to allow frames carried over the switch/router medium to be tagged, either through the use of the trunk command or using tagged hybrid interfaces. Additionally the VLAN traffic must be permitted over this link using the port trunk allow-pass vlan <vlan> or port hybrid tagged vlan <vlan> command.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# Principle and Configuration of HDLC and PPP

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- Serial has in recent years been slowly phased out in many parts of all networks in favor of Ethernet technology, however still remains active as a legacy technology in a great number of enterprise networks alongside Ethernet. Serial has traditionally provided solutions for communication over long distances and therefore remains a prominent technology for Wide Area Network (WAN) communication, for which many protocols and legacy WAN technologies remain in operation at the enterprise edge. A thorough knowledge of these technologies is required to support many aspects of WAN operation.

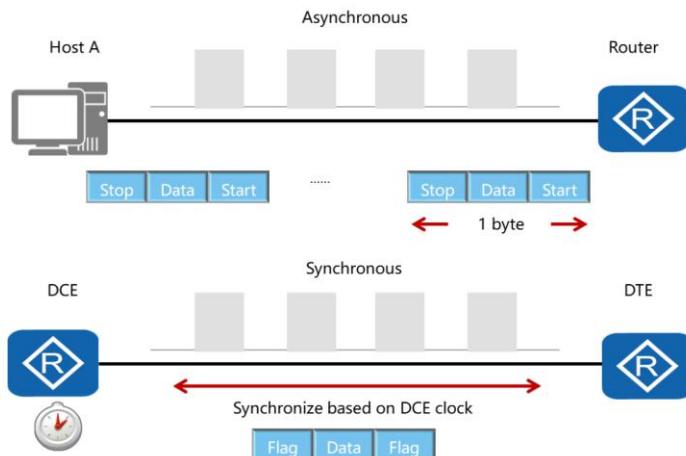


## Objectives

- Upon completion of this section, you will be able to:
  - Explain how data is carried over a serial based medium.
  - Configure link layer protocols for serial links.



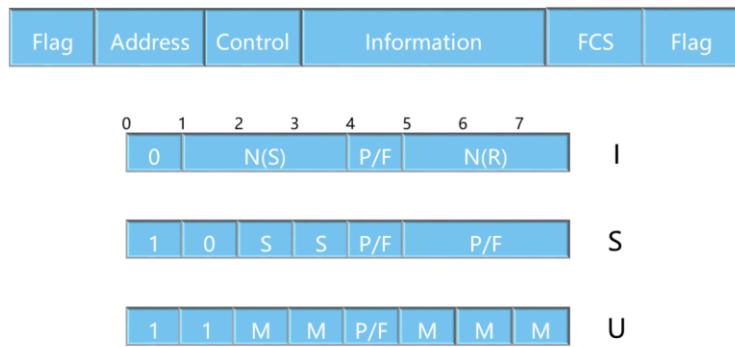
# Serial Signaling



- Serial connections represent a form of legacy technology that has commonly been used for the support of Wide Area Network (WAN) transmissions. The transmission of data as electrical signals over a serial link again requires a form of signaling to control the sending and receiving of frames as found with Ethernet. Serial connections define two forms of signaling that may be used for synchronization of transmissions, known as Asynchronous and Synchronous communication.  
Asynchronous signaling works on the principle of sending additional bits referred to as start and stop bits with each byte or frame to allow the receiving node to be aware of the incoming frame, and thus periodically reset the timing between frames to ensure that the rates between transmission and reception are maintained. The start bit is always represented as a 0 bit value while the stop bit represents a 1 bit value. One of the main concerns with this signaling method is the additional overhead as a result for each frame delivered, with the start and stop bits representing a large percentage of the frame overall. This method however is commonly associated with technologies such as Asynchronous Transfer Mode (ATM), a form of cell switching technology that generates fixed sized frames (cells) of 53 bytes as a means of supporting lower jitter through minimizing queue processing times, making it ideal for real time communication such as voice, but has begun to make way for newer technologies such as MPLS switching and due to the loss of its advantage over the frame processing speeds that are now possible with routers and switches.
- Synchronous serial connections rely on a clocking mechanism between the peering devices in which one side (DCE) provides the clocking to synchronize communication. This clocking is maintained through the carrying of clocking information between the sender and receiver as part of the data signal.



## The HDLC Protocol



- Only the information frame (I) format is used on AR2200 series.

- The High-level Data Link Control (HDLC) is a bit-oriented data link protocol that is capable of supporting both synchronous and asynchronous data transmissions. A complete HDLC frame consists of the Flag fields that are used to mark the start and end of a HDLC frame, often as 01111110, or 01111111 when a frame is to be suddenly aborted and discarded. An address field supports multipoint situations where one or multiple secondary terminals communicate with a primary terminal in a multipoint (multidrop) topology known as unbalanced connections, as opposed to the more commonly applied balanced (point to point) connections. The control field defines the frame type as either information, supervisory or unnumbered, and frame check sequence (FCS) field for ensuring the integrity of the frame.
- Of the control field frame types, only the information frame type is supported by Huawei ARG3 series routers and is used to carry data. The information frame type carries send N(S) and receive N(R) sequence numbers, as well as Poll and Final bits (P/F) for communicating status between primary and secondary stations. Supervisory frame types in HDLC are used for error and flow control and unnumbered frame types are used to manage link establishment for example between primary and secondary stations.



## Basic Configuration of HDLC

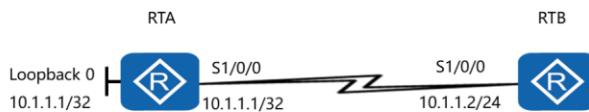


```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol hdlc
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address 10.0.1.1 30
```

- Establishment of HDLC as the link layer protocol over serial connections requires simply that the link protocol be assigned using the link-protocol hdlc command under the interface view for the serial interface that is set to use the protocol. The configuration of the link protocol must be performed on both peering interfaces that are connected to the point-to-point network before communication can be achieved.



## Assigning Unnumbered Addresses in HDLC



```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol hdlc
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address unnumbered interface loopBack 0
```

- IP addresses can be borrowed from another interface in order to establish connectivity over the serial link.

- When an interface has no IP address, it cannot generate routes or forward packets. The IP address unnumbered mechanism allows an interface without an IP address to borrow an IP address from another interface. The IP address unnumbered mechanism effectively enables the conservation of IP addresses, and does not require that an interface occupy an exclusive IP address all of the time. It is recommended that the interface that is assigned as the interface from which the unnumbered IP address is borrowed be a loopback interface since this type of interface is more likely to be always active and as such supply an available address.
- When using an unnumbered address, a static route or dynamic routing protocol should be configured so that the interface borrowing the IP address can generate a route between the devices. If a dynamic routing protocol is used, the length of the learned route mask must be longer than that of the lender's IP address mask, because ARG3 series routers use the longest match rule when searching for routes. If a static route is used and the IP address of the lender uses a 32-bit mask, the length of the static route mask must be shorter than 32 bits. If a static route is used and the IP address of the lender uses a mask less than 32 bits, the length of the static route mask must be longer than that of the lender's IP address mask.



## Configuration Validation

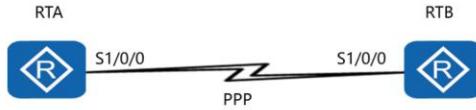
```
[RTA]display ip interface brief
*down: administratively down ^down: standby (l): loopback
(s): spoofing
.....
Interface          IP Address/Mask   Physical   Protocol
LoopBack0          10.1.1.1/32       up         up(s)
Serial1/0/0         10.1.1.1/32       up         up
Serial1/0/1         unassigned       up         down
```

- The IP address is shown to have been borrowed from the loopback interface and assigned to interface serial 1/0/0.

- Through the display ip interface brief command, a summary of the address assignment is output. In the event of assigning an unnumbered address, the address value will display as being present on multiple interfaces, showing that the IP address has been successfully borrowed from the logical loopback interface for use on the physical serial interface.



## PPP Protocol Application



- A multiprotocol standard used as with HDLC to define the link layer operation over a serial medium.

- The Point-to-Point Protocol (PPP) is a data link layer protocol that encapsulates and transmits network layer packets over point-to-point (P2P) links. PPP supports point-to-point data transmission over full-duplex synchronous and asynchronous links.
- PPP is built upon the Serial Line Internet Protocol (SLIP). PPP supports both synchronous and asynchronous links, whereas other data link layer protocols such as Frame Relay (FR) support only synchronous links. PPP is an extensible protocol, facilitating the extension of not only IP but also other protocols and is capable of supporting the negotiation of link layer attributes. PPP supports multiple Network Control Protocols (NCP) such as the IP Control Protocol (IPCP) and Internetwork Packet Exchange Control Protocol (IPXCP) to negotiate the different network layer attributes. PPP provides the Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP) for network security authentication. PPP has no retransmission mechanism, reducing the network cost and speeding up packet transmission.



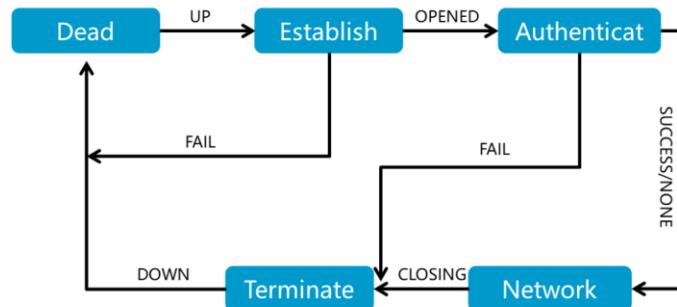
## Components of PPP

Name	Function
PPP Encapsulation Method	Defines the format to be used when supporting encapsulation of upper layer protocols such as IP, IPX etc.
Link Control Protocol	Defines the method of establishing, configuring, and testing the data-link connection.
Network Control Protocol	Defines a set of protocols for establishing a connection and negotiating parameters for different network-layer protocols.

- PPP encapsulation provides for multiplexing of different network-layer protocols simultaneously over the same link however in today's networks, the capability of PPP requires generally an IP only solution. The versatility of PPP to accommodate a variety of environments is well supported through Link Control Protocol (LCP). In order to establish communications over a point-to-point link, each end of the PPP link must first send LCP packets to configure and test the data link. More specifically LCP is used to negotiate and establish agreement for encapsulation format options, manage the MRU of packets, detect a looped-back link through magic numbers and determine errors in terms of parameter misconfigurations, as well as terminate an established link. Peer authentication on the link, and determination of when a link is functioning properly and when it is failing represent other optional facilities that are provided by LCP.
- After the link has been established and optional facilities have been negotiated as required by the LCP component of PPP, NCP packets must then be sent to choose and configure one or more network-layer protocols. Typical IP based Network Control Protocols enable features such as address configuration (IPCP), and (van Jacobson) compressed TCP/IP.



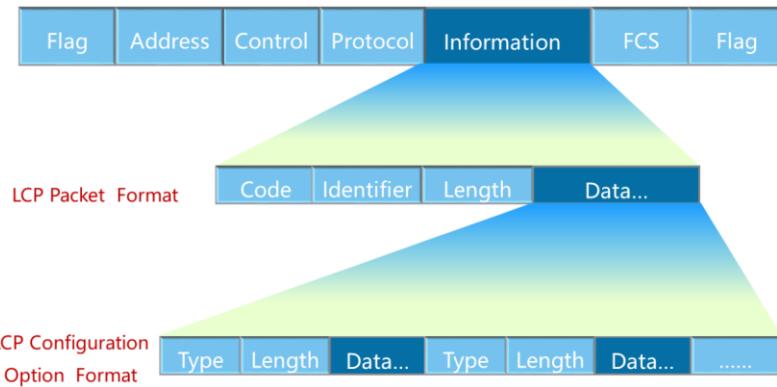
## PPP Link Establishment Process



- This initiation and termination of a PPP link begins and ends with the Dead phase. When two communicating devices detect that the physical link between them is activated (for example, carrier signals are detected on the physical link), PPP will transition from the Dead phase into the Establish phase. In the Establish phase, the two devices perform an LCP negotiation to negotiate the working mode as either single-link (SP) or multi-link (MP), the Maximum Receive Unit (MRU), authentication mode etc.
- If the authentication mode is defined, the optional Authenticate phase will be initiated. PPP provides two password authentication modes: PAP authentication and CHAP authentication. Two CHAP authentication modes are available: unidirectional CHAP authentication and bidirectional CHAP authentication. In unidirectional CHAP authentication, the device on one end functions as the authenticating device, and the device on the other end functions as the authenticated device. In bidirectional CHAP authentication, each device functions as both the authenticating device and authenticated device. In practice however, only unidirectional CHAP authentication is used. Following successful authentication, the Network phase initiates, through which NCP negotiation is performed to select and configure a network protocol and to negotiate network-layer.
- Parameters. Each NCP may be in an Opened or Closed state at any time. After an NCP enters the Opened state, network-layer data can be transmitted over the PPP link.
- PPP can terminate a link at any time. A link can be terminated manually by an administrator, or be terminated due to the loss of carrier, an authentication failure, or other causes.



## PPP Frame



- PPP generally adopts a HDLC like frame architecture for the transmission over serial connections. Flag fields are adopted to denote the start and the end of a PPP frame which is identifiable from the binary sequence 01111110 (0x7E). The address field, although present, is not applied to PPP as is the case with HDLC and therefore must always contain a 11111111 (0xFF) value, which represents an 'All-Stations' address. The control field is also fixed with a value of 00000011 (0x03) representing the unnumbered information command.
- The frame check sequence (FCS) is generally a 16 bit value used to maintain the integrity of the PPP frame. PPP additionally defines a 8 or 16 bit protocol field that identifies the datagram encapsulated in the Information field of the packet. Typical examples may include 0xc021 for Link Control Protocol, 0xc023 for Password Authentication Protocol, and 0xc223 for the Challenge Handshake Authentication Protocol. The Information field contains the datagram for the protocol specified in the Protocol field.
- The maximum length for the Information field, (not including the Protocol field), is defined by the Maximum Receive Unit (MRU), which defaults to 1500 bytes. Where the value 0xc021 is implemented, communicating devices negotiate by exchanging LCP packets to establish a PPP link.
- The LCP packet format carries a code type field that references various packet types during PPP negotiation, for which common examples include Configure-Request (0x01), Configure-Ack (0x02), Terminate-Request (0x05) etc. The Data field carries various supporting type/length/value (TLV) options for negotiation, including MRU, authentication protocols etc.



## Packet Types Used in LCP Negotiation

Packet Type	Function
Configure-Request	Include the parameters for link establishment and link configuration.
Configure-Ack	Confirmation sent once all Configure-Request parameters have been validated.
Configure-Nak	The parameters included in Configure-Request are recognized but not all accepted.
Configure-Reject	The parameters included in Configure-Request from the peer are not all recognized.

- As part of the LCP negotiation, a number of packet types are defined that enable parameters to be agreed upon before a PPP data link is established. It is necessary that the two communicating devices negotiate the link layer attributes such as the MRU and authentication mode. In order to achieve this, various packet types are communicated.
- The Configure-Request packet type allows initiation of LCP negotiation between peering devices and must be transmitted at such times. Any Configure-Request packet type sent must be responded to, and may be done so through one of a number of response packet types. Where every configuration option received in a Configure-Request is recognizable and all values are acceptable, a Configure-Ack packet type will be transmitted.
- Where all received configuration options in the Configure-Request packet type are recognized, but some values are not accepted, a Configure-Nak packet type will be transmitted, and contain only the unaccepted configuration options originally received in the Configure-Request packet type. A Configure-Reject is used when certain configuration options received in a Configure-Request are not recognizable, and thus are not accepted for negotiation.



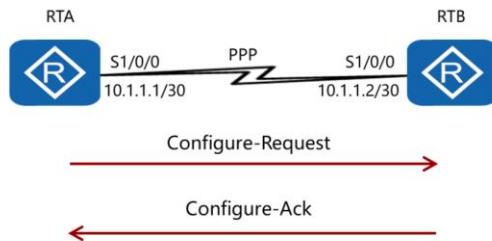
## Common Link Parameters of LCP Negotiation

Parameter	Function	Default
Maximum Receive Unit	The total length of the Information and Padding field for the PPP frame.	1500
Authentication protocol	Authentication protocol used by the peer.	No Authentication
Magic-Number	Magic-Number is generated randomly, used for link loop detection.	Enable

- Some of the common configuration options that are negotiated and carried as part of the LCP packet include the MRU, Authentication protocol supported by the sending peer as well as the magic number.
- The magic number provides a method to detect looped-back links and other anomalies at the data link layer. In the case where a Configure-Request is received containing a Magic-Number as a configuration option, the received Magic-Number is used to compare multiple received Configure-Request messages sent to the peer by comparison of the Magic-Number. If the two Magic-Numbers of the received Configure-Request messages are different, then the link is understood to be a non-looped-back link for which a Request-Ack can be given in response. If the two Magic-Numbers are equal however, then a possibility exists that the link is looped-back and that further checking must be performed for this Configure-Request, and is done so by sending a Configure-Nak to effectively request a different Magic-Number value.



## LCP Link Parameters Negotiation

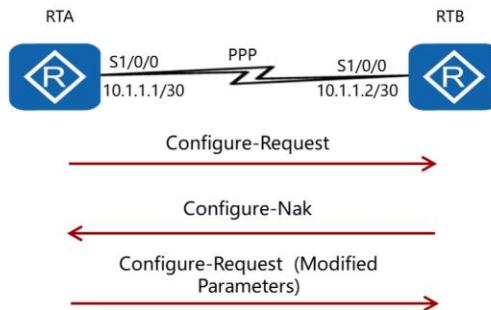


- Successful PPP negotiations result in a Configure-Ack reply to a Configure-Request packet.

- The sequence of events leading to the establishment of PPP between two peers is initiated by the sending of a Configure-Request packet to the peering device. Upon receiving this packet, the receiver must assess the configuration options to determine the packet format to respond with. In the event that all configuration options received are acceptable and recognized, the receiver will reply with a Configure-Ack packet.



## LCP Link Parameters Negotiation

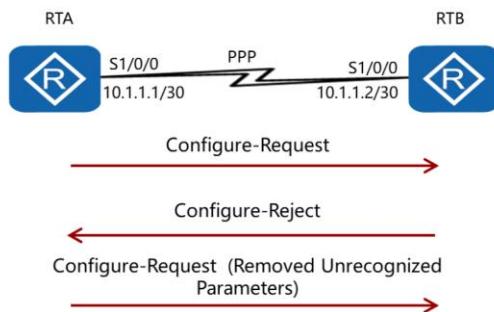


- Configure-Nak packets are generated where parameters are recognized but not all are accepted.

- Following the initial transmission of the Configure-Request as part of PPP negotiation, it is also possible that a Configure-Nak be returned, in particular where all configuration options are recognized, but some values are not accepted. On reception of the Configure-Nak packet a new Configure-Request is generated and sent, however the configuration options may generally be modified to the specifications in the received Configure-Nak packet.
- Multiple instances of a configuration option may be specified by the Configure-Nak packet for which the peer is expected to select a single value to include in the next Configure-Request packet.



## LCP Link Parameters Negotiation

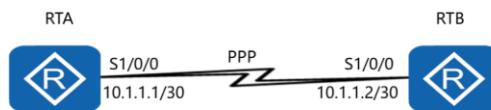


- Configure-Reject packets are generated where not all parameters are recognized by the peer.

- For PPP LCP negotiation in which one or multiple configuration options received in a Configure-Request are unrecognized or considered not acceptable for negotiation, a Configure-Reject packet is transmitted. Reception of a valid Configure-Reject indicates that when a new Configure-Request be sent, and any configuration options that are carried together with the Configure-Reject packet must be removed from the configuration options to be sent as part of the following Configure-Request packet.



## PPP Basic Configuration

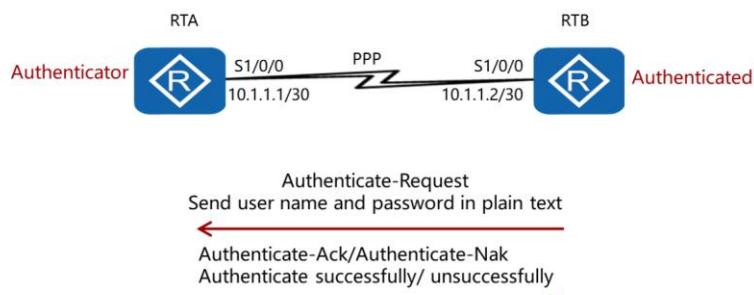


```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol ppp
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address 10.0.1.1 30
```

- Establishment of PPP requires that the link layer protocol on the serial interface be specified. For ARG3 series of routers, PPP is enabled by default on the serial interface. In the event where the interface is currently not supporting PPP, the link-protocol ppp command is used to enable PPP at the data link layer. Confirmation of the change of encapsulation protocol will be prompted, for which approval should be given as demonstrated in the configuration example.



## PPP Authentication Mode - PAP

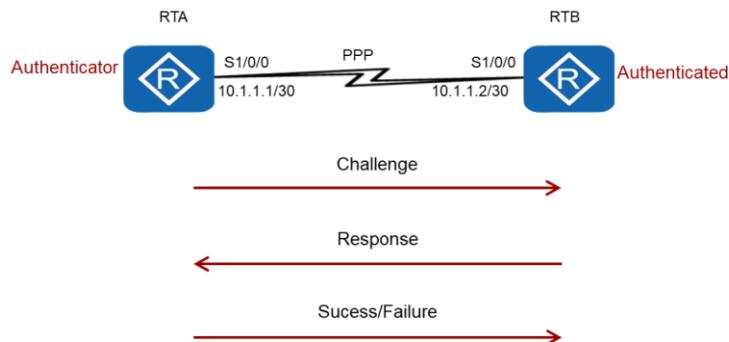


- The Password Authentication Protocol relies on the transmission of a password over the link for peer authentication.

- The Password Authentication Protocol (PAP) is a two-way handshake authentication protocol that transmits passwords in plain text. PAP authentication is performed during initial link establishment. After the Link Establishment phase is complete, the user name and password are repeatedly sent by the peer to the authenticator until authentication is acknowledged or the connection is terminated. PAP authentication effectively simulates login operations in which plain text passwords are used to establish access to a remote host. The authenticated device sends the local user name and password to the authenticator. The authenticator checks the user name and password of the authenticated device against a local user table and sends an appropriate response to the authenticated device to confirm or reject authentication.



## PPP Authentication Mode - CHAP

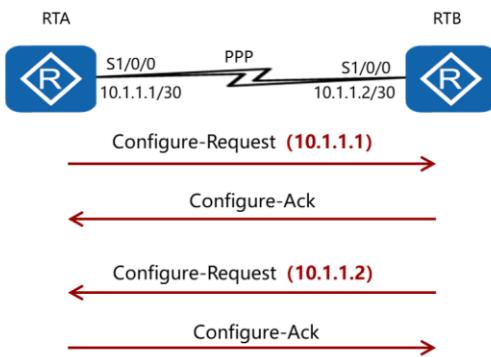


- The Challenge Handshake Authentication Protocol relies on a challenge and challenge response for peer authentication.

- The Challenge Handshake Authentication Protocol (CHAP), is used to periodically verify the identity of the peer using a three-way handshake. This is done upon initial link establishment, and can be repeated periodically. The distinguishing principle of CHAP lies in the protection given through avoiding transmission of any password over the link, instead relying on a challenge and response process that can only be successful if both authenticator and authenticated devices are supporting a value referred to as a secret. An algorithm such as MD5 is commonly used to hash any challenge and response, to ensure the integrity of the value and the resulting hash value, and is compared to a result generated by the authenticator. If both the response and value that is created by the authenticator match, the authenticated peer is approved.



## IPCP Static Address Negotiation

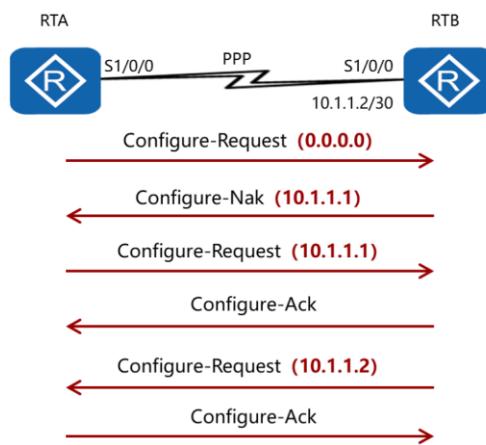


- The Internet Protocol Control Protocol (IPCP) is the Network Control Protocol (NCP) used for establishing and configuring IP.

- The IP Control Protocol (IPCP) is responsible for configuring, enabling, and disabling the IP protocol modules on both ends of the point-to-point link. IPCP uses the same packet exchange mechanism as the Link Control Protocol (LCP). IPCP packets may not be exchanged until PPP has reached the Network phase. IPCP packets received before this phase is reached are expected to be silently discarded. The address negotiation configuration option provides a way to negotiate the IP address to be used on the local end of the link, for which a statically defined method allows the sender of the Configure-Request to state which IP-address is desired. Upon configuration of the IP address a Configure-Request message is sent containing the IP address requested to be used, followed by a Configure-Ack from the peering device to affirm that the IP address is accepted.



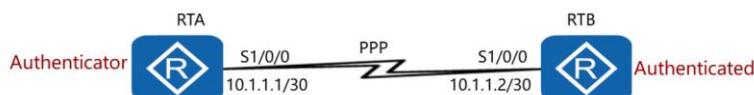
# IPCP Dynamic Address Negotiation



- A local device operating as a client and needing to be assigned an IP address in the range of the remote device (server) must make a request for a valid address by applying the ip address-ppp negotiate command on the physical interface with which the client peers with the server. Through this method, a client can retrieve a valid address. This is applicable in scenarios such as where a client accesses the Internet through an Internet Service Provider (ISP) network, and through which it can obtain an IP address from the ISP. An address is proposed to the client upon receiving a configure request for which no IP address has been defined. The PPP server (RTB) will respond with a Configure-Nak which contains suggested IP address parameters for RTA. A follow up Configure-Request message with a change to the IP addressing enables the (NCP) IPCP to successfully establish network layer protocols.



## Configuring PAP Authentication



```
[RTA] aaa  
[RTA-aaa] local-user huawei password cipher huawei123  
[RTA-aaa] local-user huawei service-type ppp  
[RTA] interface Serial 1/0/0  
[RTA-Serial1/0/0] link-protocol ppp  
[RTA-Serial1/0/0] ppp authentication-mode pap  
[RTA-Serial1/0/0] ip address 10.1.1.1 30  
  
[RTB] interface Serial 1/0/0  
[RTB-Serial1/0/0] link-protocol ppp  
[RTB-Serial1/0/0] ppp pap local-user huawei password cipher huawei123  
[RTB-Serial1/0/0] ip address 10.1.1.2 30
```

- The establishment of PAP authentication requires that one peer operate as the authenticator in order to authenticate an authenticated peer. The PPP PAP authenticator is expected to define the authentication mode, a local user name and password, and the service type. If a domain is defined to which the local user belongs (as defined by AAA), the authentication domain is also expected to be specified under the PAP authentication mode.
- An authenticated peer requires that an authentication user name, and authentication password be specified in relation to the username and password set by the authenticator. The `ppp pap local-user <username> password { cipher | simple } <password>` command is configured on the authenticated device to achieve this.



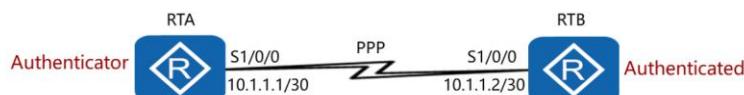
## PAP Configuration Validation

```
<RTB>debugging ppp pap all
Aug 20 2013 04:50:24.280.4+00:00 RTB PPP/7/debug2:
    PPP State Change:
        Serial1/0/0 PAP : Initial --> SendRequest
Aug 20 2013 04:50:24.290.3+00:00 RTB PPP/7/debug2:
    PPP State Change:
        Serial1/0/0 PAP : SendRequest --> ClientSuccess
....
```

- Through debugging commands which provide a real-time output of events in relation to specific protocols, the authentication request process can be viewed. As displayed in the example, a PAP authentication request is performed to which authentication establishment is deemed successful.



# Configuring CHAP Authentication



```
[RTA]aaa  
[RTA-aaa]local-user huawei password cipher huawei123  
[RTA-aaa]local-user huawei service-type ppp  
[RTA]interface Serial 1/0/0  
[RTA-Serial1/0/0]link-protocol ppp  
[RTA-Serial1/0/0]ppp authentication-mode chap  
  
[RTB]interface Serial 1/0/0  
[RTB-Serial1/0/0]link-protocol ppp  
[RTB-Serial1/0/0]ppp chap user huawei  
[RTB-Serial1/0/0]ppp chap password cipher huawei123
```

- In CHAP authentication, the authenticated device sends only the user name to the authenticating device. CHAP is understood to feature higher security since passwords are not transmitted over the link, instead relying on hashed values to provide challenges to the authenticated device based on the configured password value on both peering devices. In its simplest form, CHAP may be implemented based on local user assignments as with PAP, or may involve more stringent forms of authentication and accounting achieved through AAA and authentication/accounting servers.
- As demonstrated, the configuration of CHAP based on locally defined users requires limited configuration of local user parameters and the enablement of PPP CHAP authentication mode on the authenticator device. Where domains exist, the authenticator may also be required to define the domain being used if different from the default domain.



## CHAP Configuration Validation

```
<RTB>debugging ppp chap all
Aug 20 2013 05:15:54.230.1+00:00 RTB PPP/7/debug2:
PPP State Change:
    Serial1/0/0 CHAP : Initial --> ListenChallenge
Aug 20 2013 05:15:54.230.7+00:00 RTB PPP/7/debug2:
    PPP State Change:
        Serial1/0/0 CHAP : ListenChallenge --> SendResponse
Aug 20 2013 05:15:54.250.3+00:00 RTB PPP/7/debug2:
    PPP State Change:
        Serial1/0/0 CHAP : SendResponse --> ClientSuccess
....
```

- Debugging of the CHAP authentication processes displays the stages involved with CHAP authentication, originating from listening on the interface for any challenges being received following the LCP negotiation. In the event that a challenge is sent, the authenticated device must provide a response for which a hash value is generated, involving the set authentication parameters on the authenticated peer (password), that the authenticator will promptly validate and provide a success or failure response to.



## Summary

- Following a Configure-Request, what packet type is expected to be received before the PPP link layer can be successfully established?
- Which protocol is used to perform the negotiation of IP addresses, and during which phase is this negotiated?

- A Configure-Ack packet is required in order to allow the link layer to be successfully established when using PPP as the link layer encapsulation mode.
- The Internet Protocol Control Protocol (IPCP) is used to negotiate the IP protocol modules as part of the NCP negotiation process. This occurs during the Network phase of PPP establishment.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# Principle and Configuration of PPPoE

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The application of DSL technology relies strongly on the existing telephone infrastructure that is found in almost every household and office globally. With the continued development of newer DSL standards allowing rates of up to 100Mbps, the application of DSL as a WAN technology for home and enterprise remains firmly valid. Traditional DSL connections were established over legacy ATM networks, however Ethernet has continued to emerge as the underlying technology on which many service providers establish their networks, and therefore knowledge of PPPoE technologies remains valued for establishing DSL connectivity at the enterprise edge.

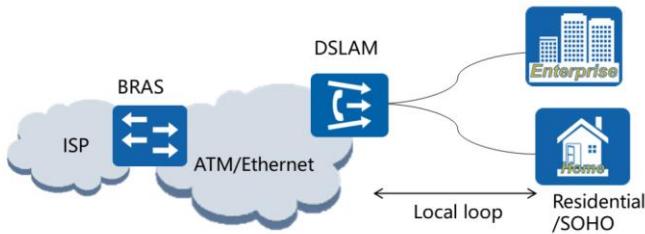


## Objectives

- Upon completion of this section, you will be able to:
  - Describe the PPPoE connection establishment process.
  - Configure a PPPoE session.



## Digital Subscriber Lines

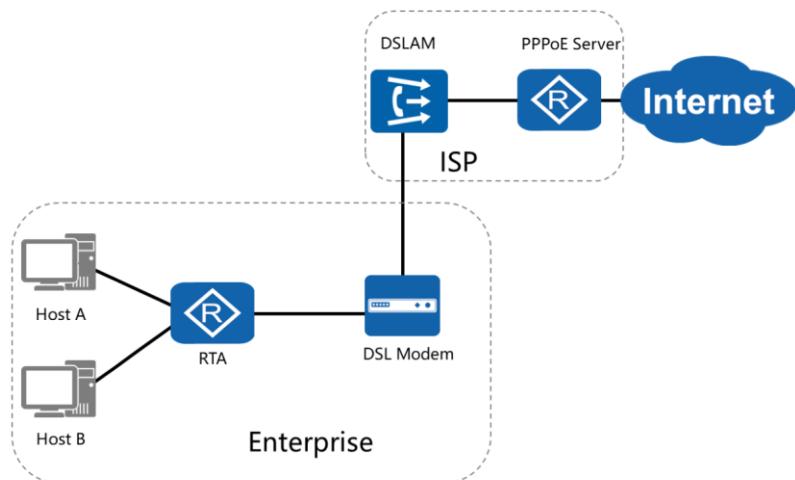


- Successive broadband technology following dial-up.
- Data signals carried over copper telephone lines, or “local loop” .

- DSL represents a form of broadband technology that utilizes existing telephony networks to allow for data communications. Communication is facilitated through a remote transceiver unit, or modem at the customer premises, which communicates over the existing telephone lines to a central office transceiver unit that takes the form of the Digital Subscriber Line Access Multiplexer (DSLAM) where traffic is multiplexed onto a high speed ATM or Ethernet network before reaching the broadband remote access server (BRAS) or PPPoA/PPPoE server within the service provider network.
- The distance between the two transceivers can vary depending on the specific DSL technology applied. In the case of an Asynchronous Digital Subscriber Line (ADSL) the distance expands up to around 18000 feet or 5,460 meters traditionally over an ATM network, whereas with a Very High Speed Digital Subscriber Line (VDSL2), local loop distances of only around 1500 meters are supported with fiber (FTTx) technology applied to provide Ethernet based backend transmission to the BRAS (PPPoE server).



## PPPoE Application in DSL





## PPPoE Protocol Packets

Type	Description
PADI	PPPoE Active Discovery Initiation (PADI) packet
PADO	PPPoE Active Discovery Offer (PADO) packet
PADR	PPPoE Active Directory Request (PADR) packet
PADS	PPPoE Active Discovery Session-Confirmation (PADS) packet
PADT	PPPoE Active Discovery Terminate (PADT) packet

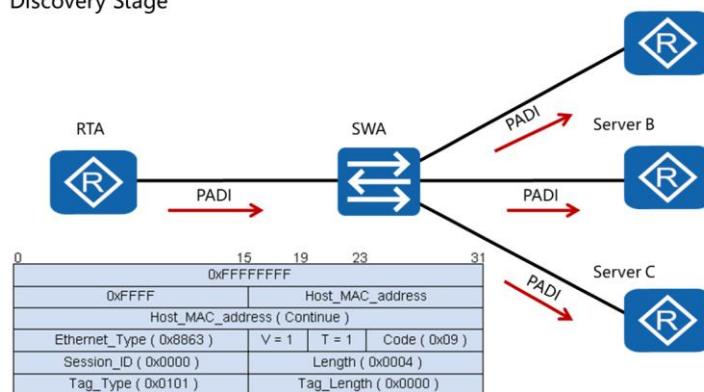
- Five packet types establish and terminate PPPoE sessions.

- PPPoE has two distinct stages, initially there is a Discovery stage followed by a PPP Session stage. When a client wishes to initiate a PPPoE session, it must first perform Discovery for which four stages are involved and rely on four individual packet types for the discovery process. These include the PPPoE Active Discovery Initiation (PADI), PPPoE Active Discovery Offer (PADO), PPPoE Active Discovery Request (PADR) and PPPoE Active Discovery Session-confirmation (PADS) protocol packet types. The termination process of the PPPoE session utilizes a PPPoE Active Discovery Terminate (PADT) packet for closure of the PPPoE session.



# PPPoE Session Establishment Process

Discovery Stage



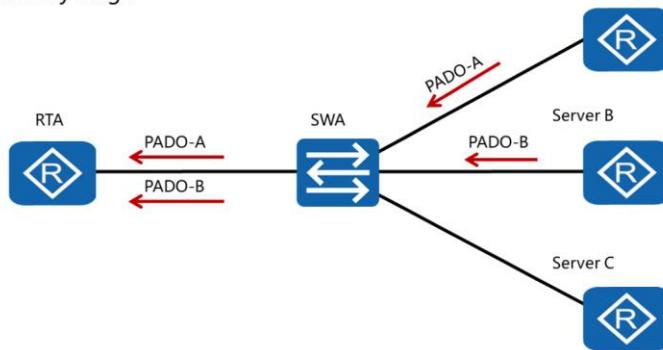
- An initiation packet is broadcast to discover access servers.

- In the Discovery stage, when a client (RTA) accesses a server using PPPoE, the client is required to identify the Ethernet MAC address of the server and establish a PPPoE session ID. When the Discovery stage is complete, both peers know the PPPoE Session\_ID and the peer Ethernet address. The PPPoE Session\_ID and peer Ethernet address define the unique PPPoE session. The client will broadcast a PPPoE Active Discovery Initiation (PADI) packet. This packet contains the service information required by the client. After receiving this PADI packet, all servers in range compare the requested services to the services that they can provide. In the PADI packet, the Destination\_address is a broadcast address, the Code field is set to 0x09, and the Session\_ID field is set to 0x0000.



## PPPoE Session Establishment Process

Discovery Stage



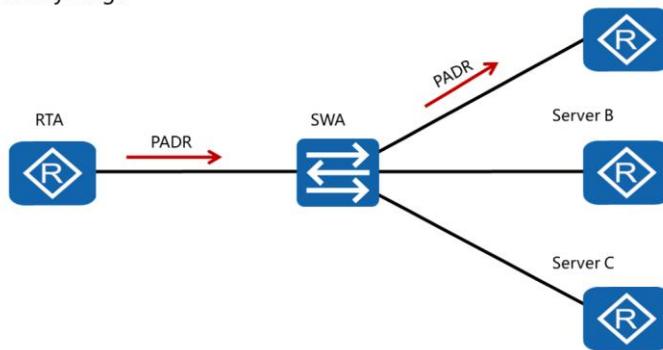
- Offers are returned to the sender by all servers that can service the received PADI packet.

- The servers that can provide the requested services send back PPPoE Active Discovery Offer (PADO) packets. The client (RTA) can receive more than one PADO packet from servers. The destination address is the unicast address of the client that sent the PADI packet. The Code field is set to 0x07 and the Session\_ID field must be set to 0x0000.



## PPPoE Session Establishment Process

Discovery Stage



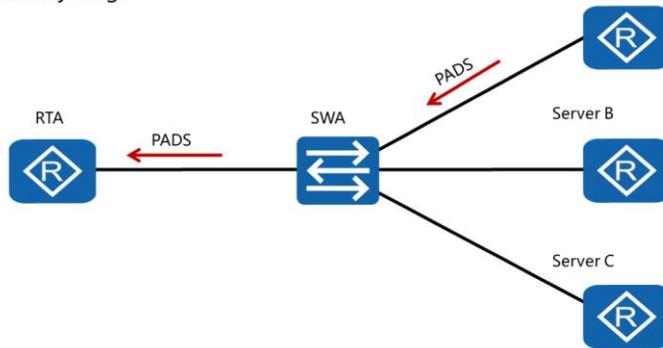
- A client responds to a chosen server based on the name or services that are provided by that server.

- Since the PADI packet is broadcast, the client can receive more than one PADO packet. It looks through all the received PADO packets and chooses one based on the AC-Name (which stands for the Access Concentrator name, and generally refers to a value that uniquely distinguishes one server from another), or the services offered by the PADO packet. The client looks through the PADO packets to choose a server, and sends a PPPoE Active Discovery Request (PADR) packet to the chosen server. The destination address is set to the unicast address of the access server that sent the selected PADO packet. The code field is set to 0x19 and the session ID field is set to 0x0000.



## PPPoE Session Establishment Process

Discovery Stage

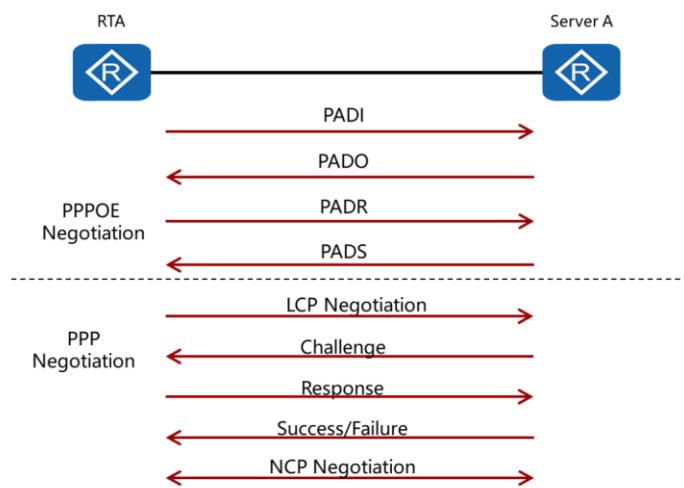


- The chosen server generates a unique PPPoE session ID in preparation for the negotiation of the PPP session.

- When receiving a PADR packet, the access server prepares to begin a PPPoE session. It generates a unique session ID for the PPPoE session and replies to the client with a PPPoE Active Discovery Session-confirmation (PADS) packet. The destination address field is the unicast Ethernet address of the client that sends the PADR packet. The code field is set to 0x65 and the session ID must be set to the value created for this PPPoE session. The server generates a unique session identifier to identify the PPPoE session with the client and sends this session identifier to the client with the PADS packet. If no errors occur, both the server and client enter the PPPoE Session stage.



## PPPoE Session Establishment Process



- Once the PPPoE session begins, PPP data is sent as in any other PPP encapsulation. All Ethernet packets are unicast. The ETHER\_TYPE field is set to 0x8864. The PPPoE code must be set to 0x00. The SESSION\_ID must not change for that PPPoE session and must be the value assigned in the discovery stage. The PPPoE payload contains a PPP frame. The frame begins with the PPP Protocol-ID.



## Packet Size Negotiation

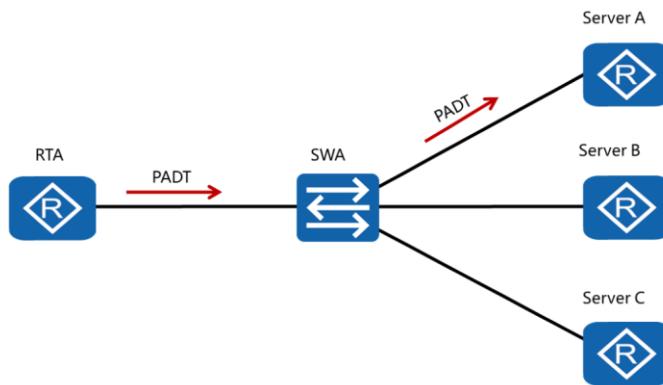


- An additional six bytes PPPoE header is carried in the frame.
- The MTU/MRU must support a lower value to prevent frame loss.

- In the PPPoE session establishment process, PPP LCP negotiation is performed, at which point the maximum receive unit (MRU) is negotiated. Ethernet typically has a maximum payload size of 1500 bytes, where the PPPoE header is 6 bytes, and the PPP Protocol ID is 2 bytes, the PPP maximum receive unit (MRU) cannot be greater than 1492 bytes since this will likely cause packet fragmentation. When LCP terminates, the client and access concentrator (server) stop using that PPPoE session. If the client needs to start another PPP session, it returns to the Discovery stage.



## PPPoE Session Termination



- Used to notify of the termination of a PPPoE session.



# Configuring a PPP Dialer Interface

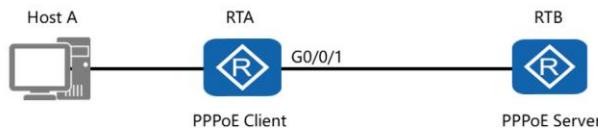


```
[RTA]dialer-rule
[RTA-dialer-rule]dialer-rule 1 ip permit
[RTA-dialer-rule]quit
[RTA]interface dialer 1
[RTA-Dialer1]dialer user enterprise
[RTA-Dialer1]dialer-group 1
[RTA-Dialer1]dialer bundle 1
[RTA-Dialer1] ppp chap user enterprise@huawei
[RTA-Dialer1] ppp chap password cipher huawei123
[RTA-Dialer1]ip address ppp-negotiate
```

- Three individual steps are required in the configuration of a PPPoE client, beginning with the configuration of a dialer interface. This enables the connection to be established on demand and a session connection to be disconnected after remaining idle for a period of time. The dialer-rule command enables the dialer-rule view to be accessed from which the dialer-rule can be set to define conditions for initiating the PPPoE connection.
- The dialer user chooses a dialer interface on which to receive a call according to the remote user name negotiated by PPP. The dialer user enables the dial control center and indicates the remote end user name, which must be the same as the PPP user name on the remote end server. If the user name parameter is not specified when using the undo dialer user command, the dial control centre function is disabled and all of the remote user names are deleted. If this parameter is specified, only the specified user name is deleted. The dialer bundle <number> command in the AR2200E is used to bind physical interfaces and dialer interfaces.
- PPP authentication parameters are defined for authentication of the client connection to the server along with the ip address ppp-negotiate command for negotiation on an interface, to allow the interface to obtain an IP address from the remote device.



## PPPoE Session Binding



```
[RTA] interface GigabitEthernet 0/0/1
[RTA-GigabitEthernet0/0/1]pppoe-client dial-bundle-number 1 on-demand
[RTA-GigabitEthernet0/0/1]quit
[RTA] ip route-static 0.0.0.0 0 dialer 1
```

- A binding is performed of the PPPoE session with the dialer bundle, and associated with the PPPoE WAN interface.

- The second step involves the establishment of the PPPoE session binding of the dialer bundle to the interface over which negotiation is to take place for the configured dialer parameters. The pppoe-client dial-bundle-number <number> command is used to achieve this where the number refers to the bundle number configured as part of the dialer parameters.
- If on-demand is not specified, the PPPoE session works in permanent online mode. If this parameter is specified, the PPPoE session works in on-demand dialing mode. The AR2200 supports the packet triggering mode for on-demand dialing. In permanent online mode, the AR2200 initiates a PPPoE session immediately after the physical link comes up. The PPPoE session persists until the undo pppoe-client dial-bundle-number command is used to delete it. In triggered online mode, the AR2200 does not initiate a PPPoE session immediately after the physical link comes up. Instead, the AR2200 initiates a PPPoE session only when data needs to be transmitted on the link. If no data is transmitted on the PPPoE link within the interval specified by dialer timer idle <seconds>, the AR2200 terminates the PPPoE session. When data needs to be transmitted on the PPPoE link, the AR2200 sets up the PPPoE session again.
- The final step requires that a default static route be configured to allow any traffic for which a network destination is not defined as a longer match in the routing table to initiate the PPPoE session through the dialer interface.



## Dialer Interface Configuration Validation

```
<Huawei>display interface Dialer 1
Dialer1 current state: UP
Line protocol current state: UP (spoofing)
Description: HUAWEI, AR Series, Dialer1 Interface
Route Port, The Maximum Transmit Unit is 1500, Hold timer is 10(sec)
Internet Address is negotiated, 192.168.10.254/32
Link layer protocol is PPP
LCP initial
Physical is Dialer
Bound to Dialer1:0:
Dialer1:0 current state : UP
Line protocol current state : UP

Link layer protocol is PPP
LCP opened, IPCP opened
```

- The display interface dialer command is used to verify the current status of the dialer configuration and allows for the locating of faults on a dialer interface. The dialer current state of UP identifies that the physical status of the interface is active, while a DOWN state would confirm that a fault currently exists. The line protocol state refers to the link layer protocol status for which an UP status confirms the active connection.
- A link that has no IP address assigned to the interface or is suppressed will show to be in a DOWN state, where suppression can be identified by dampening to the interface mainly due to a persistent interface flapping. The hold timers represents the heartbeat of the PPP dialer connection after PPP LCP negotiation transitions to Opened. The Internet address is negotiated from the IP pool that exists within the PPPoE server, where no address is present, the system will display “Internet protocol processing: disabled.” The IP address is assigned when a connection is negotiated i.e. through a ping request in the pool range.
- The LCP negotiation status defines the current state, where “initial” indicates that the physical layer is faulty or no negotiation has been started. An “Opened” state indicates that the system has sent a Configure-Ack packet to the peer device and has received a Configure-Ack packet from the peer device, as such confirming that the link layer is operating correctly.



## PPPoE Session Validation

```
[RTA] display pppoe-client session summary  
PPPoE Client Session:  
ID   Bundle   Dialer   Intf       Client-MAC     Server-MAC     State  
0     1         1         GE0/0/1      54899876830c  000000000000  IDLE
```

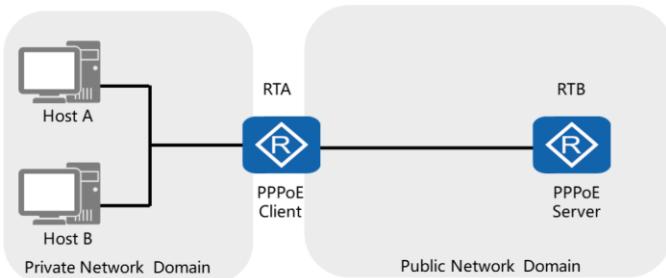
```
[RTA] display pppoe-client session summary  
PPPoE Client Session:  
ID   Bundle   Dialer   Intf       Client-MAC     Server-MAC     State  
1     1         1         GE0/0/1      00e0fc0308f6  00e0fc036781  UP
```

- The PPPoE client session status can be determined as either IDLE, in the discovery stage (PADI/PADR), or UP.

- The display pppoe-client session summary command provides information regarding the PPPoE sessions on the Point-to-Point Protocol over Ethernet (PPPoE) client, including the session status and statistics. Two examples are given to highlight the difference between the PPPoE session states. The ID represents the PPPoE ID while the bundle ID and Dialer ID are related to the values in the dialer parameter configuration.
- The interface defines the interface on which the client side negotiation is performed. The MAC address of the PPPoE client and server are also defined, however the Server-MAC is only determined after negotiation is established. The PPPoE has four states that are possible. An IDLE state indicates that the PPPoE session is currently not established and no attempt to establish the PPPoE connection has been made.
- If the state is PADI, this indicates that the PPPoE session is at the Discovery stage and a PPPoE Active Discovery Initiation (PADI) packet has been sent. A PADR state indicates that the PPPoE session is at the Discovery stage and a PPPoE Active Discovery Request (PADR) packet has been sent. Finally, an UP state indicates that the PPPoE session has been established successfully.



## PPPoE Application in Enterprise Networks



- Privately addressed hosts cannot exist in the public domain.
- Address translation along with PPPoE necessary.

- Much of the implementation of PPPoE has focused on a general point-to-point connection reflective of a typical lab environment however in real world applications the PPPoE client often represents the gateway between an enterprise network and the Wide Area Network (WAN), to which external destinations and resources are accessed.
- The internal network of an enterprise commonly employs a private network scheme in order to conserve addresses and these addresses cannot be used to establish IP communication over the public network infrastructure. This requires that the AR2200 provide the network address translation (NAT) function to translate private IP addresses into public IP addresses, and is a feature that is commonly applied to facilitate internal network communications over PPPoE.



## Summary

- Why is it necessary to reduce the MTU/MRU size of PPPoE packets?
- What is the purpose of the dialer bundle command when establishing the PPPoE connection?

- IP packets have a maximum payload size of 1500 bytes, however when used together with PPPoE, it requires an additional 6 bytes, as well as another 2 bytes for the PPP protocol ID, that causes the size of the packet to exceed the maximum supported payload size. As such the LCP negotiated MRU of a packet supporting PPPoE must be no larger than 1492 bytes.
- The dialer bundle command binds the physical interface to the dialer interface that is used to establish the PPPoE connection.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## Network Address Translation

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The continued growth of IP networks in general has resulted in an ever increasing pressure on the IPv4 address space, and the need for a way to prolong the depletion until long term solutions are founded. Network Address Translation has become well established as the existing solution and widely implemented within enterprise networks. Many variations of NAT have been developed thus conserving the public address space whilst enabling continued public network communication. This section introduces the concept of NAT along with examples of common NAT methods applied, for maintaining internetworking between the enterprise network and the public network domain.

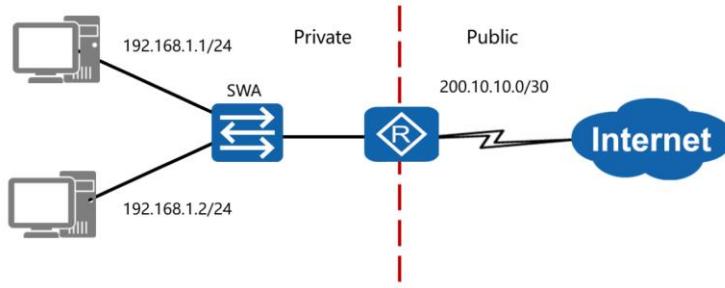


## Objectives

- Upon completion of this section, you will be able to:
  - List some of the different forms of Network Address Translation.
  - Explain the general behavior of NAT.
  - Configure NAT to suit application requirements.



## Private & Public Networks

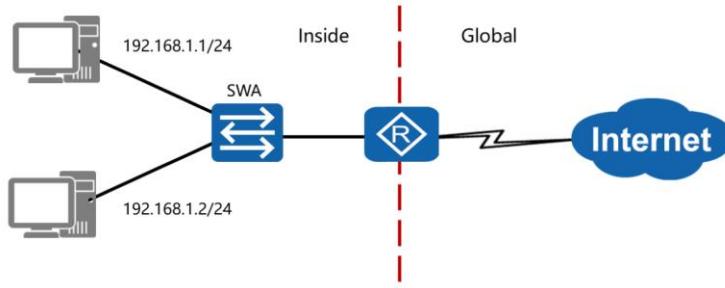


- A measure taken against rapid depletion of IP addresses.
- Gateway operates as a private/public address boundary.

- One of the main issues that has faced the expanding network has been the progressive depletion of IP addresses as a result of growing demand. The existing IPv4 addressing scheme has struggled to keep up with the constant growth in the number of devices that continue to make up the public IP network which is commonly recognized as the Internet. IPv4 addressing has already faced depletion from IANA, the industry body that is responsible for the allocation of addressing globally.
- One makeshift solution was to allocate a range of private addresses based on the existing IP address classes that could be reused. This solution has allowed network domains to implement these addressing schemes based on the private address range, and in relation to the scale of the network domain. This allows for traffic that originates and is destined for locations in the same domain to communicate without consumption of valued public addresses.
- A problem arises however on facilitating communication beyond the private network domain where destinations for traffic exist within the public domain or in another private domain beyond that public domain. Network Address Translation has become the standard solution to this issue allowing end stations to forward traffic via the public network domain from a private network.



## NAT Behavior

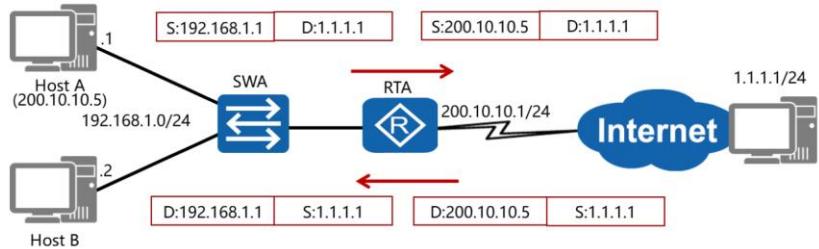


- NAT boundaries are represented as either inside or global.
- Translation of addresses is performed between boundaries.

- Network Address Translation (NAT) uses the generally established boundary of the gateway router to identify network domains for translation. Domains are considered to be either internal private networks or external public networks between which NAT is performed. The main principle lies in the reception of traffic with a source address that is in the private network and a destination address which represents a location beyond the private network domain.
- The router is expected to implement NAT to translate the private address to a public address to allow the public destination address to receive a valid return public address via which packets received can be replied. NAT must also create a mapping table within the gateway to allow the gateway to determine as to which private network destination address a packet received from the public network should be sent, again requiring address translation to be performed along the return path.



## Static NAT

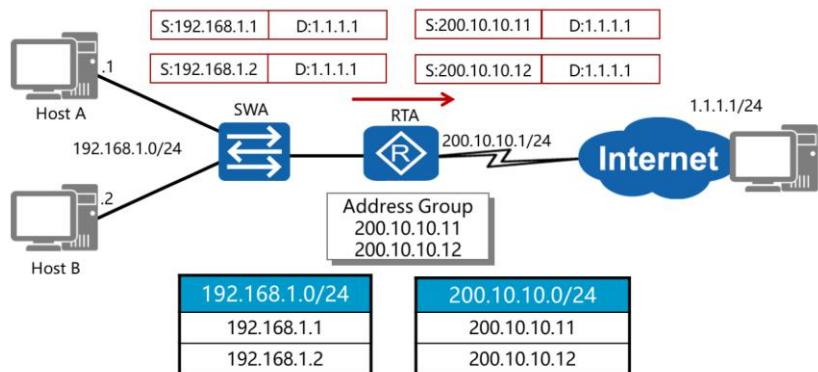


- One-to-one mapping of private to public addresses.
- Limits the need for address management with session flows.

- A number of implementations of NAT are possible and are capable of being applied to a variety of different situations. Static NAT represents a direct one-to-one mapping that allows the IP address of a specific end system to be translated to a specific public address. On a large scale the one-to-one mapping of static NAT does not do anything to alleviate the address shortage, however is applicable in cases such as where a host may wish to have certain privileges associated with an address to which the host is statically associated. This same principle may also apply to servers that may wish to be reached from the external network by a specific public address.
- In the example given packets originating from source 192.168.1.1 are destined for the public network address of 1.1.1.1. The network gateway RTA, builds a mapping between the private address of 192.168.1.1 and a public address of 200.10.10.5 that is assigned as the source address of the packet before being forwarded by the gateway to the intended destination. Any return packet will be sent as a reply with the destination address of 200.10.10.5 to which the gateway will receive and perform the static translation, before forwarding the packet to the associated host of 192.168.1.1. The static mapping of addresses requires no real management of address allocation for users, since addressing is manually assigned.



## Dynamic NAT

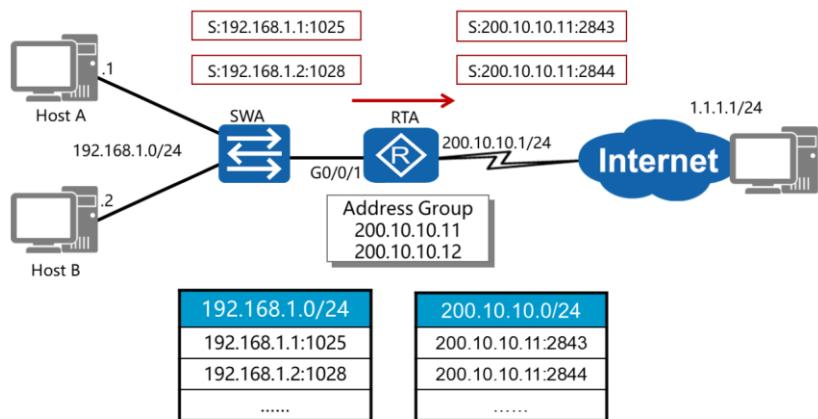


- Private address mapping based on an address resource pool.
- Allows users to utilize public addresses based on need.

- Dynamic NAT works on the principle of address pools by which internal end systems wishing to forward traffic across the public network are capable of associating with a public address from an address pool. End systems requiring to communicate with destinations in the public network domain must associate with a unique public address that is attainable from the public address range of the pool.
- An address is assigned from the NAT server address pool as each end system attempts to forward traffic to a public network destination. The number of IP addresses owned by the NAT server is far less than the number of internal hosts because not all the internal hosts access external networks at the same time. This is usually determined according to the number of internal hosts that access external networks at peak hours.
- The example demonstrates a case where two internal hosts generate packets intended for the destination 1.1.1.1/24, in which each internal host is allocated a unique address from the address pool, to allow each host to be distinguished in the public network from the other. Once communication is no longer required over the public network, the address mapping will be removed to allow the public address to be returned to the address pool.



## Network Address Port Translation

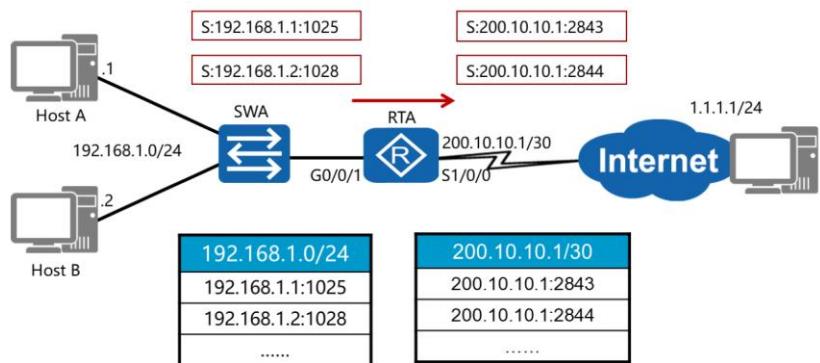


- Port numbers distinguish mapping of the same public address.

- In addition to the many-to-many address translation found within Dynamic NAT, Network Address Port Translation (NAPT) can be used to implement concurrent address translation. NAPT allows multiple internal addresses to be mapped to the same public address. It is also called many-to-one address translation or address multiplexing. NAPT maps IP addresses and interfaces. The datagrams from different internal addresses are mapped to interfaces with the same public address and different port numbers, that is, the datagrams share the same public address.
- The router receives a request packet sent from the host on the private network for accessing the server on the public network. The packet's source IP address is 192.168.1.1, and its port number is 1025. The router selects an idle public IP address and an idle port number from the IP address pool, and sets up forward and reverse NAPT entries that specify the mapping between the source IP address and port number of the packet and the public IP address and port number. The router translates the packet's source IP address and port number to the public IP address and port number based on the forward NAPT entry, and sends the packet to the server on the public network. After the translation, the packet's source IP address is 200.10.10.11, and its port number is 2843.



## Easy IP

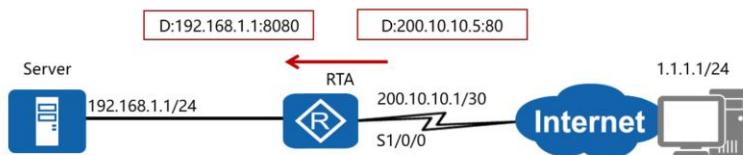


- The WAN interface address used as a single public address for all internal users, with port numbers used to distinguish sessions.

- Easy IP is applied where hosts on small-scale local area networks require access to the public network or Internet. Small-scale LANs are usually deployed where only a few internal hosts are used and the outbound interface obtains a temporary public IP address through dial-up. The temporary public IP address is used by the internal hosts to access the Internet. Easy IP allows the hosts to access the Internet using this temporary public address.
- The example demonstrates the Easy IP process. The router receives a request packet sent from the host on the private network for accessing a server in the public network. The packet's source IP address in this case is 192.168.1.1, and its port number is 1025. The router sets up forward and reverse Easy IP entries that specify the mapping between the source IP address and port number of the packet and the public IP address and port number of the interface connected to the public network. The router translates the source IP address and port number of the packet to the public IP address and port number, and sends the packet to the server on the public network. After the translation, the packet's source IP address is 200.10.10.1, and its port number is 2843.
- After receiving a response from the server, the router queries the reverse Easy IP entry based on the packet's destination IP address and port number. The router translates the packet's destination IP address and port number to the private IP address and port number of the host on the private network, and sends the packet to the host. After the translation, the packet's destination IP address is 192.168.1.1, and its port number is 1025.



## NAT Internal Server

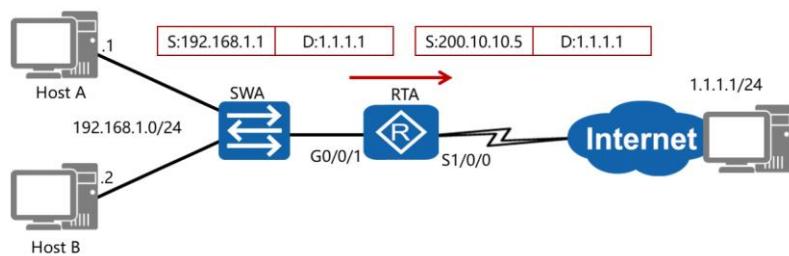


- External sources can reach internal addresses.
- Mapping of both the IP address and port number is performed.

- NAT can shield hosts on private networks from public network users. When a private network needs to provide services such as web and FTP services for public network users, servers on the private network must be accessible to public network users at any time.
- The NAT server can address the preceding problem by translating the public IP address and port number to the private IP address and port number based on the preconfigured mapping.
- Address translation entries of the NAT server are configured on the router, after which the router receives an access request sent from a host on the public network. The router queries the address translation entry based on the packet's destination IP address and port number. The router translates the packet's destination IP address and port number to the private IP address and port number based on the address translation entry, and sends the packet to the server on the private network. The destination IP address of the packet sent by the host on the public network is 200.10.10.5, and the destination port number of 80. After translation is performed by the router, the destination IP address of the packet is 192.168.1.1, and its port number is 8080. After receiving a response packet sent from the server on the private network, the router queries the address translation entry based on the packet's source IP address and port number. The router translates the packet's source IP address and port number to the public IP address and port number based on the address translation entry, and sends the packet to the host on the public network. The source of the response packet sent from the host on the private network is 192.168.1.1, and its port number is 8080. After translation by the router, the source IP address of the packet is 200.10.10.5, and the port number is again port 80.



## Static NAT Configuration



```
[RTA] interface GigabitEthernet0/0/1  
[RTA-GigabitEthernet0/0/1] ip address 192.168.1.254 24  
[RTA] interface Serial1/0/0  
[RTA-Serial1/0/0] ip address 200.10.10.1 24  
[RTA] nat static global 200.10.10.5 inside 192.168.1.1
```

- Static NAT indicates that a private address is statically bound to a public address when NAT is performed. The public IP address in static NAT is only used for translation of the unique and fixed private IP address of a host. The `nat static [protocol {<tcp>}|<udp>}global { <global-address>}| current-interface <global-port>} inside {<host-address> <host-port>} vpn-instance <vpn-instance-name> netmask <mask> acl <acl-number> description <description>]` command is used to create the static NAT and define the parameters for the translation.
- The key parameters applied as in the example are the global parameter to configure the external NAT information, specifically the external address, and the inside parameter that allows for the internal NAT information to be configured. In both cases the address and port number can be defined for translation. The global port specifies the port number of the service provided for external access. If this parameter is not specified, the value of global-port is 0. That is, any type of service can be provided. The host port specifies the service port number provided by the internal server. If this parameter is not specified, the value of host-port is the same as the value of global-port.



## Static NAT Configuration Validation

```
[RTA]display nat static
Static Nat Information:
Interface : Serial1/0/0
Global IP/Port      : 200.10.10.5/----
Inside IP/Port       : 192.168.1.1/----
Protocol : ----
VPN instance-name   : ----
Acl number          : ----
Netmask            : 255.255.255.255
Description         : ----

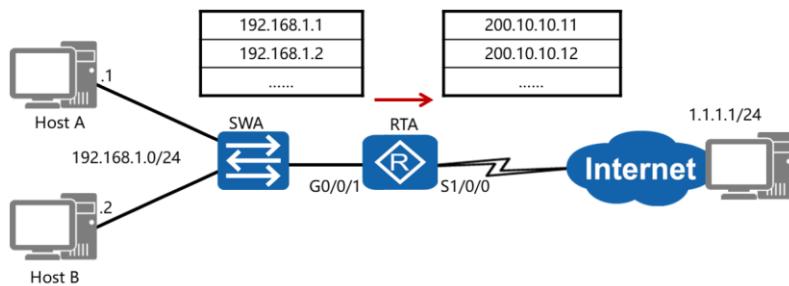
Total : 1
```

- Static inside and global address translation can be verified.

- The configuration of static NAT can be viewed through the display nat static command. The command displays the network address translation information with regards to the interface through which the address translation is performed, the global and inside addresses which are translated along with the used ports. Where the ports are not defined, a null result will be displayed. The configuration for translation may be protocol specific to either TCP or UDP protocol traffic, in which case the Protocol entry will be defined.



## Dynamic NAT Configuration



```
[RTA]nat address-group 1 200.10.10.11 200.10.10.16
[RTA]acl 2000
[RTA-acl-basic-2000]rule 5 permit source 192.168.1.0 0.0.0.255
[RTA-acl-basic-2000]quit
[RTA]interface serial1/0/0
[RTA-Serial1/0/0]nat outbound 2000 address-group 1 no-pat
```

- The configuration of dynamic network address translations involves implementation of the nat outbound command. It relies on the prior configuration of a network access control list that is used to specify a rule to identify specific traffic to which an event or operation will be applied. The details regarding access control lists are covered in a later unit. An association is made between these access control list rules and the NAT address pool. In this manner, the addresses specified in the access control list can be translated by using the NAT address pool.
- The example demonstrates how the nat outbound command has been associated with an access control list with an identifier of 2000 to allow traffic from the 192.168.1.0/24 network range to be translated as part of an address group referred to as address-group 1. This defines a pool range from 200.10.10.11 through to 200.10.10.16 which internal addresses will employ for address translation. The no-pat parameter in the command means that no port address translation will occur for the addresses in the pool, therefore each host must translate to a unique global address.



## Dynamic NAT Configuration Validation

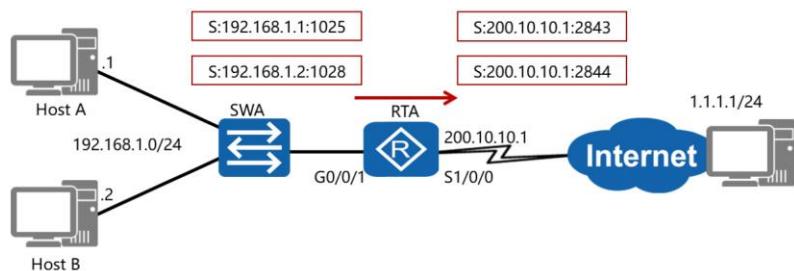
```
[RTA]display nat address-group 1
NAT Address-Group Information:
-----
Index      Start-address      End-address
1          200.10.10.11        200.10.10.16

[RTA]display nat outbound
NAT Outbound Information:
-----
Interface      Acl      Address-group/IP/Interface      Type
-----
Serial1/0/0    2000          1          no-pat
-----
Total : 1
```

- Enables group binding parameter configuration to be verified.



## Easy IP Configuration



```
[RTA]acl 2000  
[RTA-acl-basic-2000]rule 5 permit source 192.168.1.0 0.0.0.255  
[RTA-acl-basic-2000]quit  
[RTA]interface serial1/0/0  
[RTA-Serial1/0/0]nat outbound 2000
```

- The Easy IP configuration is very similar in configuration to that of dynamic network address translation, relying on the creation of an access control list rule for defining the address range to which translation is to be performed and application of the nat outbound command. The main difference is in the absence of the address group command since no address pool is used in the configuration of Easy IP. Instead the outbound interface address is used to represent the external address, in this case that being external address 200.10.10.1 of interface serial1/0/0. Additionally it is necessary that port address translation be performed, and as such the no-pat parameter cannot be implemented where an address group does not exist. The nat outbound 2000 represents a binding between the NAT operation and the access control list rule detailing the address range to which the translation will apply.



## Easy IP Configuration Validation

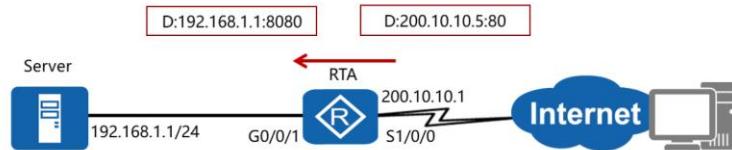
```
[RTA] display nat outbound
NAT Outbound Information:
-----
Interface      Acl      Address-group/IP/Interface      Type
-----
Serial1/0/0    2000    200.10.10.1                  easyip
-----
Total : 1
```

- Associated outbound interface parameters are displayed.
- The type field verifies the successful configuration of Easy IP.

- The same display nat outbound command can be used to observe the results of the nat outbound configuration and verify its correct implementation. The interface and access control list (ACL) binding can be determined as well as the interface (in this case) for the outbound translation. The type listing of easyip makes it clearly understood when Easy IP has been successfully configured.



## NAT Internal Server Configuration



```
[RTA]interface GigabitEthernet0/0/1  
[RTA-GigabitEthernet0/0/1]ip address 192.168.1.254 24  
[RTA]interface Serial1/0/0  
[RTA-Serial1/0/0]ip address 200.10.10.1 24  
[RTA]nat server protocol tcp global 200.10.10.5 www inside  
192.168.1.1 8080
```

- Where it is necessary to provide internal access to external users such as in the case of a public server which is part of an internal network, the NAT server configuration can be performed to enable traffic destined for an external destination address and port number to be translated to an internal destination address and port number.
- The nat server [protocol {<tcp >}|<udp>}global {<global-address> | current-interface <global port>} inside {<host-address ><host-port>} vpn-instance <vpn-instance-name> acl <acl-number> description <description >] command enables the external internal translation to be performed, where protocol identifies a specific protocol (either TCP or UDP depending on the service) to be permitted for translation along with the global address, indicating the external address for translation and the inside address relating to the internal server address.
- The port number for the external address should be defined, and commonly relates to a specific service such as http (www) on port 80. As a means of further improving the overall shielding of internal port numbers from external threats, an alternative port number can be applied to the inside network and translated through the same means as is used for address translation.



## NAT Internal Server Configuration Validation

```
[RTA]display nat server
Nat Server Information:
Interface : Serial1/0/0
    Global IP/Port      : 200.10.10.5/80 (www)
    Inside IP/Port     : 192.168.1.1/8080
    Protocol          : 6(tcp)
    VPN instance-name : -----
    Acl number        : -----
    Description       : -----

Total : 1
```

- Successful translation of the IP address and port is achieved.

- The `display nat server` command details the configuration results to verify the correct implementation of the NAT server. The interface defines the point at which the translation will occur. The global and inside IP addresses and associated port numbers can be verified. In this case the global address of 202.10.10.5 with a port number of 80 (www) will be translated to an inside server address of 192.168.1.1 with a port number of 8080 as additional security against potential port based attacks. Only TCP traffic that is destined for this address and port will be translated.



## Summary

- Which form of translation will allow a server in a DMZ to be accessed from both an external and an internal network?
- What is the function of the PAT feature?

- The NAT internal server configuration will allow a unique public address to be associated with a private network server destination, allowing inbound traffic flow from the external network after translation. Internal users are able to reach the server location based on the private address of the server.
- The PAT feature will perform translation based on the port number as well as the IP addresses. It is used as a form of address conservation where the number of public addresses available for translation are limited, or insufficient to support the number of private addresses that require possible translation.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## Access Control Lists

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- Many technologies and protocols depend on Access Control Lists (ACL) for greater management and filtering of traffic as part of security measures or application requirements. The implementation of ACL in support of other technologies, and as a form of security are required to be understood, and as such common forms of ACL solutions are introduced.

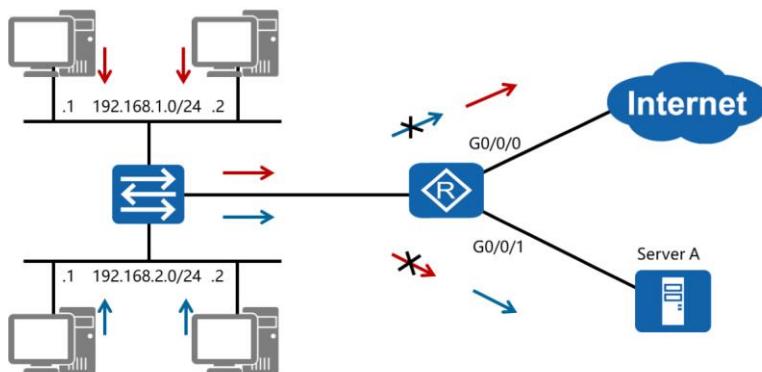


## Objectives

- Upon completion of this section, you will be able to:
  - Describe the applications for ACL in the enterprise network.
  - Explain the decision making behavior of Access Control Lists.
  - Successfully implement Basic and Advanced Access Control Lists.



## Filtering Restricted Traffic

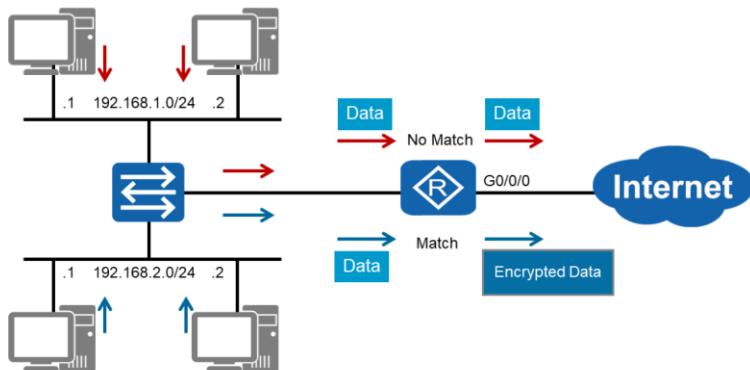


- Packets are filtered based on addresses and parameters.
- Rules allow packets to be either permitted or denied.

- An Access Control List (ACL) is a mechanism that implements access control for a system resource by listing the entities based on specific parameters that are permitted to access the resource, as well as the mode of access that is granted. In general, an ACL can be understood as a means for filtering, and can be applied to control the flow of traffic as well as identify traffic to which special operations should be performed.
- A common application involves a gateway that may have a forwarding destination to multiple networks, however may contain an ACL that manages which traffic can flow to which destination. In the example given, the network 192.168.1.0/24 is generally seen as capable of accessing the external network, in this case the Internet, whereas hosts that are represented by the 192.168.2.0/24 network are unable to forward traffic in the same manner, and therefore resulting in a transmission failure. In the case of Server A, the reverse applies with permission for access being granted by the gateway to network 192.168.2.0/24 but restricted for all hosts that are part of the 192.168.1.0/24 network.



## Filtering Interesting Traffic



- Packets can be filtered to manipulate behavior and actions.
- Parameters and forwarding behavior can be altered as a result.

- Where filtering is performed based on interesting traffic, there is no general restriction made but instead additional operations are likely to be performed which affects the current data. The example demonstrates a scenario where inbound data is filtered based on certain criteria such as in this case, the source IP address, and where an access control list is found to apply to data, associated actions are taken. Common actions may involve the changing of parameters in routed IP traffic for protocols such as the route metrics in RIP and OSPF, and also in initiating encrypted network communications for the interesting traffic, as is often applied as part of technologies such as Virtual Private Networks (VPN).



## ACL Types

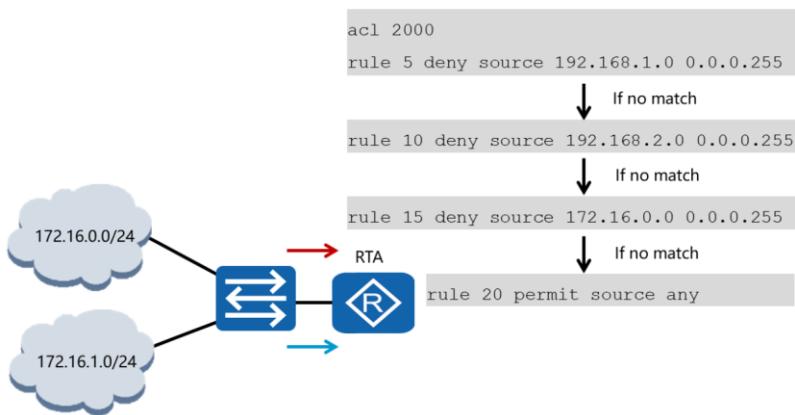
Types	Value Ranges	Parameters
Basic	2000-2999	Source IP
Advanced	3000-3999	Source & Destination IP, Protocol, Source & Destination Port
Layer 2 ACL	4000-4999	MAC Address

- Three forms of ACL can be applied to AR2200 series routers.
- Parameters for packet filtering vary for each ACL type.

- There are three general ACL types that are defined as part of the ARG3 series, including basic, advanced and layer2 access control list types. A basic ACL matches packets based on information such as source IP addresses, fragment flags, and time ranges, and is defined by a value in the range of 2000-2999. An advanced ACL provides a greater means of accuracy in parameter association, and matches packets based on information such as source and destination IP addresses, source and destination port numbers, and protocol types.
- Advanced ACL are associated with a value range from 3000-3999. Lastly is the layer 2 ACL which matches packets based on packet based Layer 2 information, such as source MAC addresses, destination MAC addresses, and Layer 2 protocol types. Traffic is filtered based on rules containing the parameters defined by each type of ACL.



## ACL Rule Management

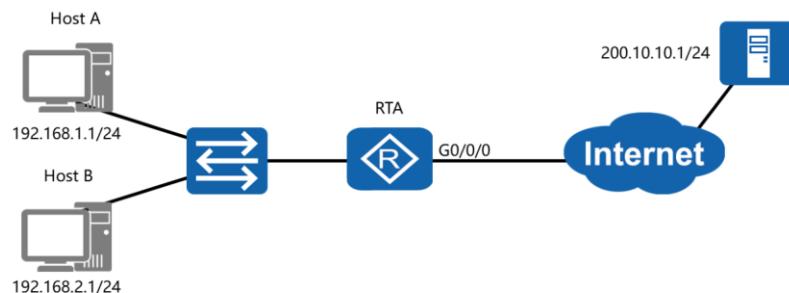


- Rules are used to manage the decision process for each ACL.

- Access Control Lists work on the principle of ordered rules. Each rule contains a permit or deny clause. These rules may overlap or conflict. One rule can contain another rule, but the two rules must be different. The AR2200 supports two types of matching order: configuration order and automatic order. The configuration order indicates that ACL rules are matched in ascending order of rule identifiers, while the automatic order follows the depth first principle that allows more accurate rules to be matched first. The configuration order is used by default and determines the priorities of the rules in an ACL based on a rule ID. Rule priorities are such able to resolve any conflict between overlapping rules. For each rule ID the ACL will determine whether the rule applies. If the rule does not apply, the next rule will be considered. Once a rule match is found, the rule action will be implemented and the ACL process will cease. If no rule matches the packet, the system does not process the packet.
- In the example, packets originating from two networks are subjected to an ACL within RTA. Packets from networks 172.16.0.0 and 172.17.0.0 will be assessed based on rule ID (configuration) order by default. Where the rule discovers a match for the network based on a wildcard mask, the rule will be applied. For network 172.16.0.0, rule 15 will match any packets with the address of 172.16.0.X where X may refer to any binary value in the octet. No specific rule matching the network 172.17.0.0 is found in the access control list and so will not be subjected to the ACL process, however in the interests of good ACL design practice, a catch all rule has been defined in rule 20 to ensure that all networks for which there is no specifically defined rule, are permitted to be forwarded.



## Basic ACL



```
[RTA]acl 2000  
[RTA-acl-basic-2000]rule deny source 192.168.1.0 0.0.0.255  
[RTA-acl-basic-2000]rule permit source 192.168.2.0 0.0.0.255  
[RTA]interface GigabitEthernet 0/0/0  
[RTA-GigabitEthernet0/0/0]traffic-filter outbound acl 2000
```

- The creation of a basic access control list requires that the administrator first identify the traffic source to which the ACL is to apply. In the example given, this refers to the source locations containing an IP address in the 192.168.1.0/24 range for which all packets containing a source IP address in this range will be discarded by the gateway. In the case of hosts that make up the 192.168.2.0/24 network range, traffic is permitted and no further action is taken for these packets. The basic ACL is applied to the interface Gigabit Ethernet 0/0/0 in the outbound direction, therefore only packets that meet both the interface and direction criteria will be subjected to ACL processing.



## Configuration Validation

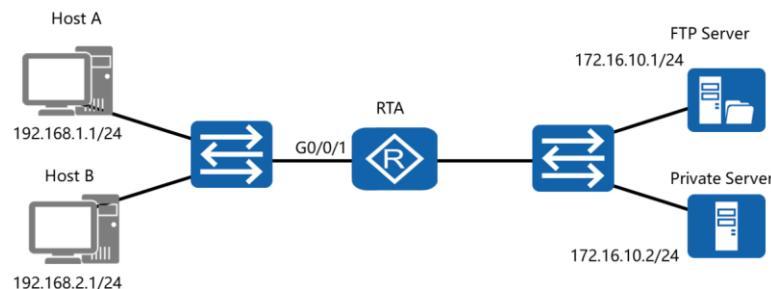
```
Host A> ping 200.10.10.1
Ping 200.10.10.1: 32 data bytes, Press Ctrl_C to break
Request timeout!
Request timeout!
Request timeout!
...
[RTA]display acl 2000
Basic ACL 2000, 2 rules
Acl's step is 5
rule 5 deny source 192.168.1.0 0.0.0.255 (5 matches)
rule 10 permit source 192.168.2.0 0.0.0.255
```

- The rules and matching order can be verified for each ACL.
- Basic ACL rules are matched based on each source IP address.

- The validation of the configured basic ACL can be achieved through the `display acl <acl-number>` where the `acl-number` refers to the basic ACL number that has been assigned to the configured ACL. The resulting output confirms the rules that have been created to deny (drop) any IP packets with the source IP address in the range 192.168.1.0/24 and permit addressing in the range 192.168.2.0/24.
- It should also be noted that each rule is automatically assigned a rule number as part of the access control list creation. The rule number defines the order in which the rules are processed and set in increments of 5 by default in Huawei ARG3 series routers. There is an ACL step between rule numbers. For example, if an ACL step is set to 5, rules are numbered 5, 10, 15, and so on. If an ACL step is set to 2 and rule numbers are configured to be automatically generated, the system automatically generates rule IDs starting from 2. The step makes it possible to add a new rule between existing rules. It is possible to configure the rule number as part of the basic ACL where required.



## Advanced ACL



```
[RTA] acl 3000
[RTA-acl-adv-3000]rule deny tcp source 192.168.1.0 0.0.0.255
destination 172.16.10.1 0.0.0.0 destination-port eq 21
[RTA-acl-adv-3000] rule deny ip source 192.168.2.0 0.0.0.255
destination 172.16.10.2 0.0.0.0
[RTA-GigabitEthernet0/0/1]traffic-filter inbound acl 3000
```

- Advanced access control lists enable filtering based on multiple parameters to support a greater detailed route selection process. While a basic ACL provides filtering based on the source IP address, Advanced ACL are capable of filtering based on the source and destination IP, source and destination port numbers, protocols of both the network and transport layer and parameters found within each layer such as IP traffic classifiers and TCP flag values (SYN|ACK|FIN etc.).
- An advanced ACL is defined by an ACL value in the range of 3000-3999 as displayed in the example, for which rules are defined to specify restriction of TCP based packets that originate from all source addresses in the range of 192.168.1.1 through to 192.168.1.255 where the destination IP is 172.16.10.1 and the destination port is port 21. A similar rule follows to define restriction of all IP based packets originating from sources in the 192.168.2.0/24 range from reaching the single destination of 172.16.10.2. A catch all rule may generally be applied to ensure that all other traffic is processed by the ACL, generally through a permit or deny statement for all IP based packets.



## Configuration Validation

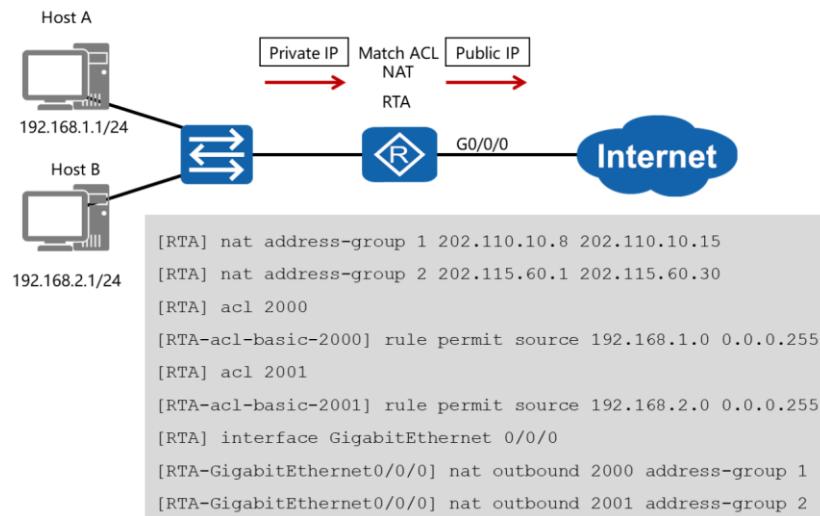
```
[RTA]display acl 3000
Advanced ACL 3000, 2 rules
Acl's step is 5
rule 5 deny tcp source 192.168.1.0 0.0.0.255 destination 172.16.10.1 0
destination-port eq ftp
rule 10 deny ip source 192.168.2.0 0.0.0.255 destination 172.16.10.2 0
```

- Advanced ACL rules defined in the range of 3000-3999 add complexity due to the number of parameters used for filtering.

- The validation of the configured advanced ACL can be achieved through the display acl <acl-number> where the acl-number refers to the advanced ACL number that has been assigned to the configured ACL. The resulting output confirms that three rules have been created to deny any TCP packets with the source IP address in the range 192.168.1.0/24 destined for 172.16.10.1 from reaching port 21(ftp), and from any source IP address in the range of 192.168.2.0/24 from reaching the destination IP address of 172.16.10.2 , while permitting all other IP traffic.



## ACL Application - NAT



- ACL can also be applied to the Network Address Translation (NAT) operation to allow filtering of hosts based on IP addresses to determine which internal networks are translated via which specific external address pools should multiple pools exist. This may occur where an enterprise network is a customer to connections from multiple service providers for which various internal users that are considered part of different networks/sub-networks wish to be translated and forwarded based on a specific address group, potentially occurring over alternative router uplink interfaces to the different service provider networks.
- In the example given, a simplified example of this is recreated where hosts within the internal network are to be filtered based on basic ACL rules for which a dynamic NAT solution is applied that allows translation to a given public address of a certain address group. In particular hosts originating from the 192.168.1.0/24 network will be translated using public addresses in the pool associated with address group 1, while hosts in the 192.168.2.0/24 network will be filtered based on the pool associated with address group 2. The *nat outbound <acl-number> address-group <address-group number>* command is implemented under the Gigabit Ethernet interface view to bind NAT in the outbound direction with the ACL, and the related IP address range is referenced by the specified address group.



## Summary

- The advanced access control list is capable of filtering traffic based on which attributes?
- Once an ACL rule is matched to a condition, what action is taken?

- Advanced ACL are capable of filtering based on the source and destination IP, source and destination port numbers, protocols of both the network and transport layer and parameters found within each layer such as IP traffic classifiers and TCP flag values (SYN|ACK|FIN etc.).
- Once a rule match is found in the access control list to a tested condition, the rule action will be implemented and the remaining ACL process will not continue.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# AAA

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- AAA defines a security architecture that is comprised of three functions referred to as Authentication, Authorization and Accounting. Each of these functions represents a modular component which can be applied as components of the security framework implemented by an enterprise, and often managed through the use of client/server based protocols such as RADIUS and HWTACACS. Implementation of the AAA architecture as a solution for enhanced functionality is introduced to reinforce the overall security framework of the enterprise network.

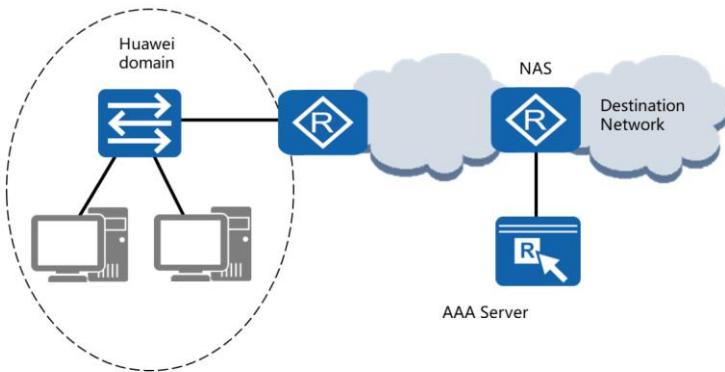


## Objectives

- Upon completion of this section, you will be able to:
  - Describe the schemes of the AAA security architecture.
  - Successfully configure Authentication and Authorization schemes.



## AAA Application

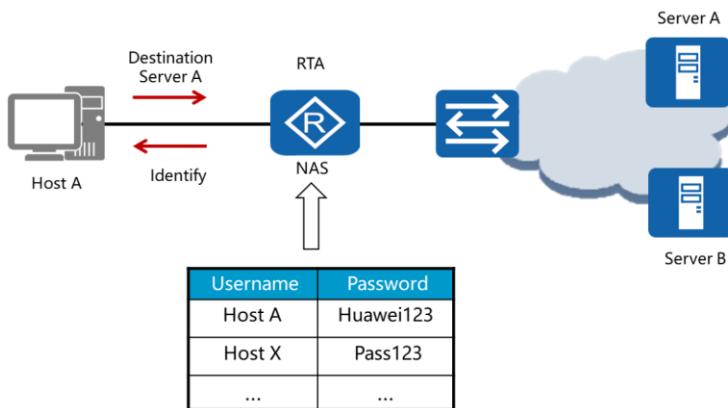


- AAA enables the authentication, authorization and accounting of users attempting to access destination network resources.

- Authentication, Authorization, and Accounting (AAA) is a technology that is used to check whether a user has permission to access a network, authorizes exactly what a user is allowed to access, and makes records regarding the network resources used by a user. VRP is capable of supporting the AAA authentication and authorization services locally within the ARG3 series of routers, which is commonly referred to as a Network Access Server or NAS, however accounting services are generally supported through an external AAA accounting server. The example here demonstrates how users that are considered part of the Huawei domain are able to gain access to resources that are located within the displayed destination network. The Network Access Server (NAS) operates as the gateway device that may perform authentication and authorization of users, or support the AAA server's authentication and authorization of users. In the case of the AAA server, those users that are authenticated and authorized to access the destination network may also initiate accounting within the AAA server for the duration of users active session.



# Authentication

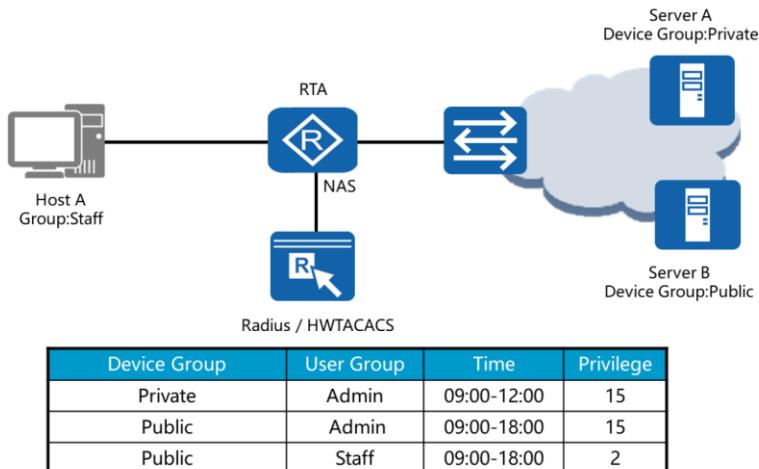


- User access is managed based on an authentication scheme.

- AAA supports three authentication modes. Non-authentication completely trusts users and does not check their validity. This is seldom used for obvious security reasons. Local authentication configures user information, including the user name, password, and attributes of local users, on a Network Access Server (NAS). Local authentication has advantages such as fast processing and low operation costs. The disadvantage of local authentication is the limited information storage because of the hardware. Remote authentication configures user information including the user name, password, and attributes on the authentication server. AAA can remotely authenticate users using the Remote Authentication Dial In User Service (RADIUS) protocol or the Huawei Terminal Access Controller Access Control System (HWTACACS) protocol. As the client, the NAS communicates with the RADIUS or HWTACACS server.
- If several authentication modes are used in an authentication scheme, these authentication modes take effect in the sequence with which the configuration modes were configured. If remote authentication was configured before local authentication and if a login account exists on the local device but is unavailable on the remote server, the AR2200 considers the user using this account as having failed to be authenticated by the remote authentication, and therefore local authentication is not performed. Local authentication would be used only when the remote authentication server did not respond. If non-authentication is configured, it must be configured as the last mode to take effect.



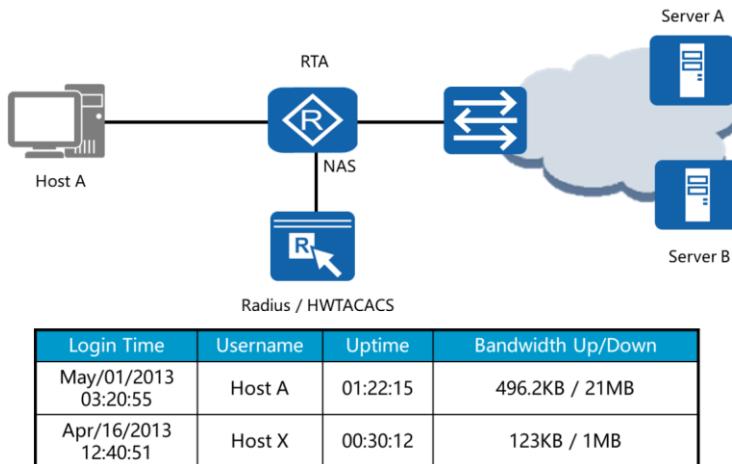
## Authorization



- The AAA Authorization function is used to determine the permission for users to gain access to specific networks or devices, as such AAA supports various authorization modes. In non-authorization mode users are not authorized. Local authorization however authorizes users according to the related attributes of the local user accounts configured on the NAS. Alternatively, HWTACACS can be used to authorize users through a TACACS server.
- An If-authenticated authorization mode can be used where users are considered authorized in the event that those users are able to be authenticated in either the local or remote authentication mode. RADIUS authorization authorizes users after they are authenticated using a RADIUS authentication server. Authentication and authorization of the RADIUS protocol are bound together, so RADIUS cannot be used to perform only authorization. If multiple authorization modes are configured in an authorization scheme, authorization is performed in the sequence in which the configuration modes were configured. If configured, non-authorization must be the last mode to take effect.



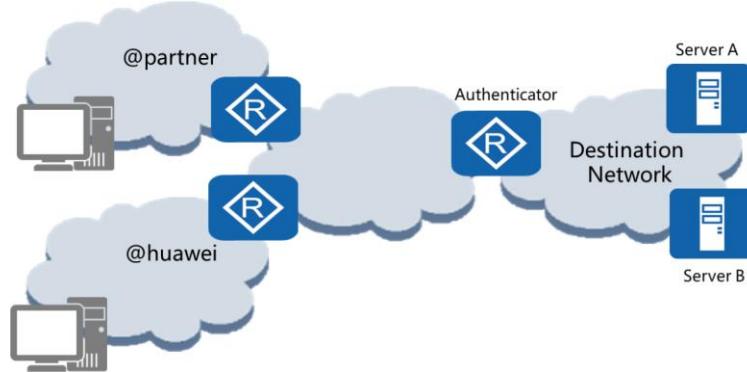
## Accounting



- The accounting process can be used to monitor the activity and usage of authorized users who have gained access to network resources. AAA accounting supports two specific accounting modes. Non-accounting can be used, and provides free services for users without any record of users, or activity logs.
- Remote accounting on the other hand supports accounting using the RADIUS server or the HWTACACS server. These servers must be used in order to support accounting due to the requirement for additional storage capacity necessary to store information regarding access and activity logs of each authorized user. The example demonstrates a very general representation of some of the typical information that is commonly recorded within user accounting logs.



## AAA Domains

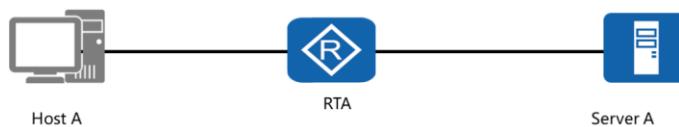


- Different schemes can be applied to users in different domains.

- The device uses domains to manage users. Authentication, authorization, and accounting schemes can be applied to a domain so that the device can authenticate, authorize, or charge users in the domain using the schemes. Each user of the device belongs to a domain. The domain to which a user belongs is determined by the character string suffixed to the domain name delimiter that can be @, |, or %.
- For example, if the user name is user@huawei, the user belongs to the huawei domain. If the user name does not contain an @, the user belongs to the default domain named default in the system. The device has two default domains: default (global default domain for common access users) and default\_admin (global default domain for administrators). The two domains can be modified but cannot be deleted.
- If the domain of an access user cannot be obtained, the default domain is used. The default domain is used for access users such as NAC access users. Local authentication is performed by default for users in this domain. The default\_admin domain is used for administrators such as the administrators who log in using HTTP, SSH, Telnet, FTP, and terminals. Local authentication is performed by default for users in this domain. The device supports a maximum of 32 domains, including the two default domains.



## AAA Local Configuration



```
[RTA] aaa
[RTA-aaa]local-user huawei password cipher hello123
[RTA-aaa]authentication-scheme auth1
[RTA-aaa-authen-auth1]authentication-mode local
[RTA-aaa-authen-auth1]quit
[RTA-aaa] authorization-scheme auth2
[RTA-aaa-author-auth2]authorization-mode local
[RTA-aaa-author-auth2]quit
[RTA-aaa]domain huawei
[RTA-aaa-domain-huawei]authentication-scheme auth1
[RTA-aaa-domain-huawei]authorization-scheme auth2
```

- Authentication and authorization can be applied on the AR2200E.

- The AR2200 router can be used as a Network Access Server (NAS) in order to implement authentication and authorization schemes. The example demonstrates the typical process that is necessary in order to successfully implement local AAA. Users for authentication must be created using the local-user <user-name> password [cipher \simple]<password> privilege level <level > command. This command specifies a user name. If the user-name contains a domain name delimiter such as @, the character before @ is the user name and the character behind @ is the domain name. If the value does not contain @, the entire character string is the user name and the domain name is the default domain.
- An authentication scheme is created in order to authenticate users and must be created before additional authentication-relevant configuration can be performed. The authentication scheme must be defined as either local, radius, hwtacacs or none. With the exception of the none parameter, the other authentication modes can be listed in the order in which authentication is to be attempted, for example should the authentication-mode hwtacacs local command be used, and if HWTACACS authentication fails, the AR2200E router will start local authentication. The authorization scheme must also be created to authorize users (except in the case of Radius server support), by creating an authorization scheme defining the authorization-mode. The authorization-mode command supports modes for hwtacacs, local, if-authenticated and none authorization.
- The domain <domain-name> command is used to create a new domain, and the implementation of the authentication-scheme <authentication-scheme> and the authorization-scheme <authorization-scheme> under the domain view will apply the authentication and authorization schemes to that domain.



## AAA Local Configuration Verification

```
[Huawei]display domain name huawei
Domain-name : huawei
Domain-state : Active
Authentication-scheme-name : auth1
Accounting-scheme-name : default
Authorization-scheme-name : auth2
Service-scheme-name : -
RADIUS-server-template : -
HWTACACS-server-template : -
User-group : -
```

- Local AAA schemes are associated with individual domains.

- The configuration of the local AAA domain and schemes for both authentication and authorization can be verified through the *display domain* or *display domain name <domain-name>* commands. Using the *display domain* command provides brief information regarding all domains that have been created, including the domain name and a domain index that is used to reference each created domain.
- The *display domain name <domain-name>* command provides specific configuration details in reference to the domain defined under the domain-name parameter. Along with the domain-name is the domain-state which presents as either Active or Block, where block refers to a domain that is in blocking state, and prevents users in this domain from being able to log in. This is implemented through an optional *state [active | block]* command implemented under the AAA domain view. A domain is in an active state after being created by default.
- The authentication-scheme-name associates the created authentication-scheme with the domain; the same applies for the authorization-scheme. The accounting-scheme is not configured locally and therefore an accounting scheme has not been associated with the created domain, and as such the default accounting scheme is listed for the domain by default. In the event that non-local (i.e RADIUS or HWTACACS) configuration is implemented to support AAA services, these will be associated with the domain under the server-template fields.



## Summary

- Which two AAA schemes are supported when configuring VRP to support the local mode?
- If no domain is defined for users, what action is taken?

- The configuration of VRP in local mode will supports both authentication and authorization schemes, the accounting scheme requires the support of remote management through a HWTACACS or RADIUS server.
- If a user is created without defining the domain to which the user belongs, the user will be automatically associated with the default domain, named default.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## Securing Data with IPSec VPN

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- Early TCP/IP protocol development did very little for ensuring the security of communications between peering devices. As networks evolved so did the need for greater protection of the data transmitted. Solutions for data protection were developed, from which IPSec emerged as a security architecture for the implementation of confidentiality, integrity and data origin authentication, primarily through the support of underlying protocols. IPSec remains a key framework in the protection of data, which has seen an integration of IPSec components adopted into the next generation of TCP/IP standards.



## Objectives

- Upon completion of this section, you will be able to:
  - Explain the basic principles of the IPSec security architecture.
  - Configure IPSec peering between two devices.



## IPSec VPN Application

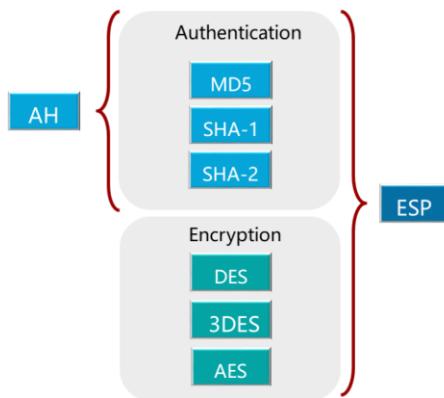


- Facilitates the establishment of private network communication over a public network infrastructure.

- Internet Protocol security (IPSec) is a protocol suite defined by the Internet Engineering Task Force (IETF) for securing Internet Protocol (IP) communication by authenticating and/or encrypting each IP packet of a communication session. Two communicating parties can encrypt data and/or authenticate the data originating at the IP layer to ensure data confidentiality, integrity and service availability.
- Confidentiality is the security service that protects the disclosure of typically application level data, but also encompasses disclosure of communication which also can be a concern in some circumstances. Traffic flow confidentiality is the service that is applied to conceal source and destination addresses, message length, or frequency of communication.
- IPSec supports two forms of integrity; connectionless and a form of partial sequence integrity. Connectionless integrity is a service that detects modification of an individual IP datagram, without regard to the ordering of the datagram in a stream of traffic. The form of partial sequence integrity offered in IPSec is referred to as anti-replay integrity, and it detects arrival of duplicate IP datagrams (within a constrained window). This is in contrast to connection-oriented integrity, which imposes more stringent sequencing requirements on traffic, e.g., to be able to detect lost or re-ordered messages. Although authentication and integrity services often are cited separately, in practice they are intimately connected and almost always offered in tandem.
- Availability, when viewed as a security service, addresses the security concerns engendered by attacks against networks that deny or degrade service. For example, in the IPSec context, the use of anti-replay mechanisms in the underlying protocols, specifically AH and ESP, support availability to prevent attacks that malicious users initiate by resending captured packets.



# IPSec VPN Architecture

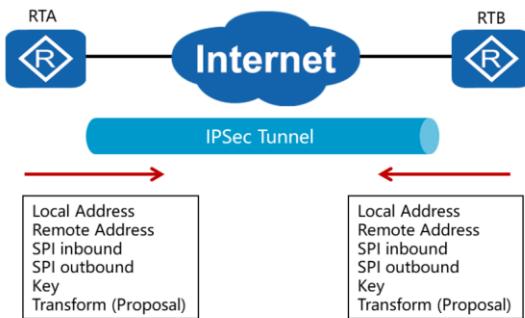


- Confidentiality and integrity of services are supported through authentication and encryption based protocols.

- IPSec uses two protocols to provide traffic security, Authentication Header (AH) and Encapsulating Security Payload (ESP). The IP Authentication Header (AH) provides connectionless integrity, data origin authentication, and an optional anti-replay service. The Encapsulating Security Payload (ESP) protocol may provide confidentiality (encryption), and limited traffic flow confidentiality.
- ESP optionally provides confidentiality for traffic. The strength of the confidentiality service depends in part on the encryption algorithm employed, for which three main forms of encryption algorithm are possible. ESP may also optionally provide authentication, however the scope of the authentication offered by ESP is narrower than for AH since the external IP header (tunnel mode) and the ESP header is/are not protected under ESP authentication.
- Data Encryption Standard with Cipher Block Chaining (DES-CBC) is a symmetric secret key algorithm that uses a key size of 64-bits, but is commonly known as a 56-bit key as the key has 56 significant bits; the least significant bit in every byte is the parity bit. Its use today is limited due to the capability for computational power to perform brute force attacks within a significantly limited time frame. Triple Data Encryption Standard (3DES) is a 168 bit encryption algorithm that involves an iteration of the DES process with the addition of an initialization vector (IV), a value used to disguise patterns in encrypted data to improve the overall confidentiality of data. The Advanced Encryption Standard (AES) supports up to 256 bit encryption employing again the CBC and IV for additional strength of encryption. As encryption becomes additionally more complex, so does the processing time needed to achieve encryption and decryption.
- AH supports an MD5 algorithm that takes as input, a message of arbitrary length and produces as output a 128-bit "fingerprint" or "message digest" of that input, while the SHA-1 produces a 160-bit message digest output. SHA-2 represents a next generation of authentication algorithms to extend security to authentication following discovery of certain weaknesses within SHA-1.
- The term SHA-2 is not officially defined but is usually used to refer to the collection of the algorithms SHA-224, SHA-256, SHA-384, and SHA-512. VRP currently supported by the AR2200E however provides support only to SHA-256, SHA-384 and SHA-512, where the value represents the bit digest length.



## Security Association

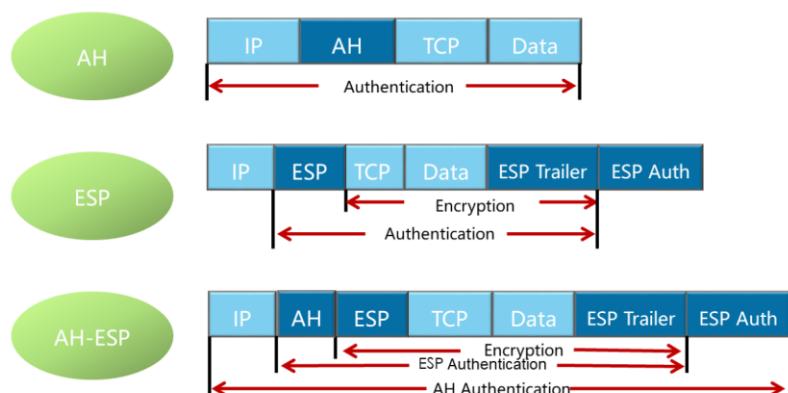


- Specifies parameters for connection establishment.
- A Security Association defines parameters in only one direction.

- A Security Association (SA) denotes a form of connection that is established in a single direction through which security services relevant to the traffic are defined. Security services are attributed to an SA in the form of either AH, or ESP, but not both. If both AH and ESP based protection is applied to a traffic stream, then two (or more) SAs are created to attribute protection to the traffic stream. In order to secure bi-directional communication between two hosts, or two security gateways as shown in the example, two Security Associations (one in each direction) are required.
- IPSec SAs are established in either a manual mode or Internet Key Exchange (IKE) negotiation mode. Establishing SAs in manual mode requires all information such as those parameters displayed in the example, be configured manually. The SAs established in manual mode however will never age. Establishing SAs using IKE negotiation mode is simpler as IKE negotiation information needs to be configured only on two peers and SAs are created and maintained by means of IKE negotiation, for which the complexity exists mainly in the IKE automated negotiation process itself, and as such will not be covered in detail here. The SA established in IKE negotiation has a time-based or traffic-based lifetime. When the specified time or traffic volume is reached, an SA becomes invalid. When the SA is about to expire, IKE will negotiate a new SA.



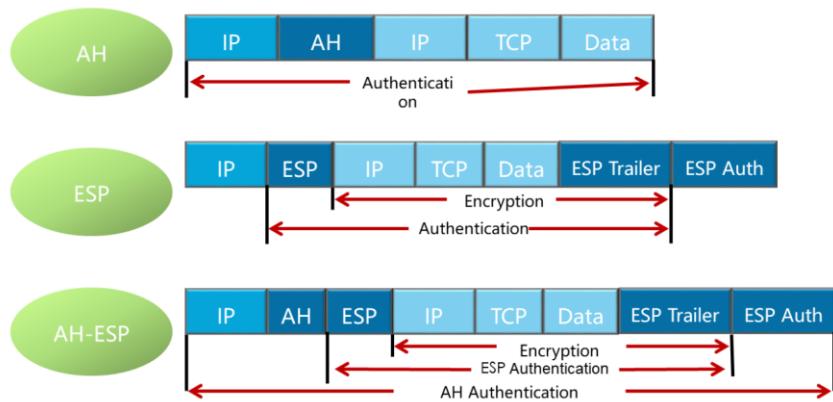
## IPSec Transport Mode



- Encapsulation modes are defined in Security Associations.
- Transport mode secures only the payload of the packet.



## IPSec Tunnel Mode

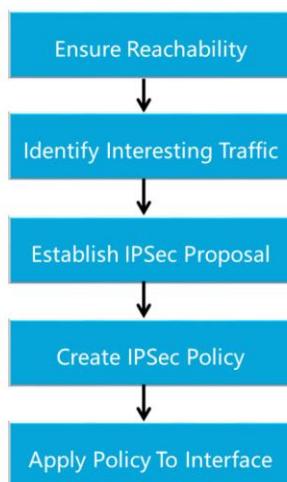


- Tunnel mode encapsulates packets in a second IP header.
- Security is extended to the inner IP header and packet payload.

- In the IPSec context, using ESP in tunnel mode, especially at a security gateway, can provide some level of traffic flow confidentiality. The outer IP header source address and destination address identify the endpoints of the tunnel. The inner IP header source and destination addresses identify the original sender and recipient of the datagram, (from the perspective of this tunnel), respectively.
- The inner IP header is not changed except to decrement the TTL, and remains unchanged during its delivery to the tunnel exit point. No change occurs to IP options or extension headers in the inner header during delivery of the encapsulated datagram through the tunnel. An AH or ESP header is inserted before the original IP header, and a new IP header is inserted before the AH or ESP header.
- The example given shows the IPSec tunnel mode during TCP based packet transmission. If AH is also applied to a packet, it is applied to the ESP header, Data Payload, ESP trailer, and ESP Authentication Data (ICV), if fields are present.



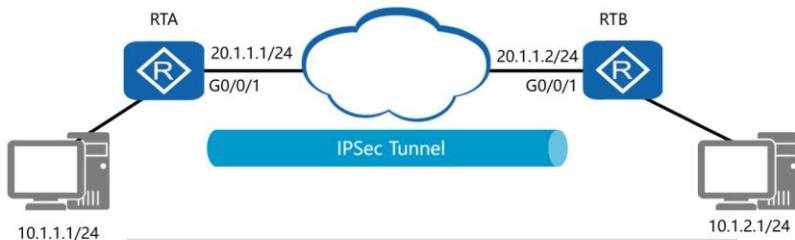
# IPSec VPN Establishment



- The implementation of an IPSec Virtual Private Network (VPN) requires that network layer reachability between the initiator and the recipient is verified to ensure non-IPSec communication is possible before building any IPSec VPN. Not all traffic is required to be subjected to rules of integrity or confidentiality, and therefore traffic filtering needs to be applied to the data flow to identify the interesting traffic for which IPSec will be responsible. A data flow is a collection of traffic and can be identified by the source address/mask, destination address/mask, protocol number, source port number, and destination port number. When using VRP, data flows are defined using ACL groups. A data flow can be a single TCP connection between two hosts or all traffic between two subnets. The first step to configure IPSec involves defining these data flows.
- An IPSec proposal defines the security protocol, authentication algorithm, encryption algorithm, and encapsulation mode for the data flows to be protected. Security protocols AH and ESP can be used independently or together. AH supports MD5, SHA-1 and SHA-2 authentication algorithms. ESP supports three authentication algorithms (MD5, SHA-1 and SHA-2) and three encryption algorithms (DES, 3DES, and AES). IPSec also supports both transport and tunnel encapsulation modes. To transmit the same data flow, peers on both ends of a security tunnel must use the same security protocol, authentication algorithm, encryption algorithm, and encapsulation mode. To implement IPSec between two gateways, it is advised to use tunnel mode to shield the actual source and destination IP addresses used in communication.
- An IPSec policy defines the security protocol, authentication algorithm, encryption algorithm, and encapsulation mode for data flows by referencing an IPSec proposal. The name and sequence number uniquely identify an IPSec policy. IPSec policies are classified into IPSec policies used for manually establishing SAs, and IPSec policies used for establishing SAs through IKE negotiation.
- To configure an IPSec policy used for manually establishing SAs, set parameters such as the key and SPI. If the tunnel mode is configured, the IP addresses for two endpoints of a security tunnel also must be set. When configuring an IPSec policy used for establishing SAs through IKE negotiation, parameters such as the key and SPI do not need to be set as they are generated through IKE negotiation automatically. IPSec policies of the same name but different sequence numbers comprise an IPSec policy group.
- In an IPSec policy group, an IPSec policy with a smaller sequence number has a higher priority. After an IPSec policy group is applied to an interface, all IPSec policies in the group are applied to the interface. This enables different SAs to be used for different data flows.



## IPSec VPN Configuration



```
[RTA] ip route-static 10.1.2.0 24 20.1.1.2
[RTA] acl number 3001
[RTA-acl-adv-3001]rule 5 permit ip source 10.1.1.0
0.0.0.255 destination 10.1.2.0 0.0.0.255
[RTA] ipsec proposal tran1
[RTA-ipsec-proposal-tran1]esp authentication-algorithm
sha1
```

- Communication at the network layer represents the initial prerequisite of any IPsec VPN connection establishment. This is supported in this example through the creation of a static route pointing to the next hop address of RTB. It is necessary to implement static routes in both direction to ensure bidirectional communication between networks. An advanced ACL is created to identify the interesting traffic for which the IPsec VPN will be initiated, whereby the advanced ACL is capable of filtering based on specific parameters and will choose to either Discard, Bypass or Protect the filtered data.
- The ipsec proposal command creates an IPSec proposal and enters the IPSec proposal view. When configuring an IPSec policy, it is necessary to reference an IPSec proposal for specifying the security protocol, encryption algorithm, authentication algorithm, and encapsulation mode used by both ends of the IPSec tunnel. A new IPSec proposal created using the ipsec proposal command by default uses the ESP protocol, DES encryption algorithm, MD5 authentication algorithm, and tunnel encapsulation mode. These can be redefined using various commands under the IPSec proposal view. The transform [ah | ah-esp | esp] will enable the current transform set to be altered, the encapsulation-mode {transport | tunnel } can alter the means used to encapsulate packets.
- The authentication algorithm used by the ESP protocol can be set using the esp authentication-algorithm [md5 | sha1 | sha2-256 | sha2-384 | sha2-512 ] and esp encryption-algorithm [des | 3des | aes-128 | aes-192 | aes-256 ] for ESP encryption assignment. The AH protocol ah authentication-algorithm [md5 | sha1 | sha2-256 | sha2-384 | sha2-512 ] command enables AH based authentication to be set.



## IPSec VPN Proposal Verification

```
[RTA]display ipsec proposal
Number of proposals : 1
IPSec proposal name : tran1
Encapsulation mode   : Tunnel
Transform             : esp-new
ESP protocol         : Authentication SHA1-HMAC-96
Encryption           DES
```

- Displays the parameters of an IPSec proposal.
- Proposal parameters must match for both peering interfaces.

- Verification of the parameters for the IPSec proposal can be achieved through the display ipsec proposal [name <proposal-name>] command. As a result it is possible to view select proposals or all proposals that have been created. The name defines the name given to the IPSec proposal created, while the encapsulation mode lists the current mode that is being used for this given proposal, which may either be transport or tunnel. Transform refers to the algorithms being applied where this may be AH, ESP or a combination of AH and ESP transforms. Depending on the transform defined, the associated algorithms will be listed.



## IPSec Policy Creation

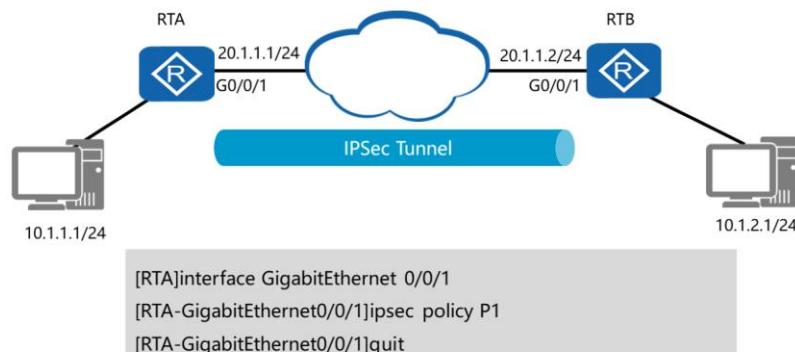
```
[RTA]ipsec policy P1 10 manual
[RTA-ipsec-policy-manual-P1-10]security acl 3001
[RTA-ipsec-policy-manual-P1-10]proposal tran1
[RTA-ipsec-policy-manual-P1-10]tunnel remote 20.1.1.2
[RTA-ipsec-policy-manual-P1-10]tunnel local 20.1.1.1
[RTA-ipsec-policy-manual-P1-10]sa spi outbound esp 54321
[RTA-ipsec-policy-manual-P1-10]sa spi inbound esp 12345
[RTA-ipsec-policy-manual-P1-10]sa string-key outbound esp simple huawei
[RTA-ipsec-policy-manual-P1-10]sa string-key inbound esp simple huawei
```

- IPSec policy defines parameters for establishing an IPSec SA.
- An IPSec policy binds the proposal parameters and traffic filters.

- The policy-name and seq-number parameters identify an IPSec policy. Multiple IPSec policies with the same IPSec policy name constitute an IPSec policy group. An IPSec policy group contains a maximum of 16 IPSec policies, and an IPSec policy with the smallest sequence number has the highest priority. After an IPSec policy group is applied to an interface, all IPSec policies in the group are applied to the interface to protect different data flows. Along with the policy-name and seq-number, the policy must define the creation of SA either manually or through IKE negotiation.
- If IKE negotiation is used, parameters must be set using the ipsec-policy-template command, while where set manually parameters are configured as shown in the example. The created ACL is bound with the policy along with the created proposal. The tunnel local and tunnel remote commands define the interface at which the IPSec tunnel will establish and terminate. A source parameter index (SPI) is defined both inbound and outbound.
- The inbound SPI on the local end must be the same as the outbound SPI on the remote end. The outbound SPI on the local end must be the same as the inbound SPI on the remote end. This number is used to reference each SA, bearing in mind that an SA applies in only one direction therefore requiring SPI in both directions to be specified. Finally authentication keys must be defined again both inbound and outbound. The outbound authentication key on the local end must be the same as the inbound authentication key on the remote end. Where IKE negotiation is used, both the SPI and authentication keys are automatically negotiated.



## Applying Policies to Interfaces



```
[RTA]interface GigabitEthernet 0/0/1  
[RTA-GigabitEthernet0/0/1]ipsec policy P1  
[RTA-GigabitEthernet0/0/1]quit
```

- The IPSec policy is bound to the physical interface via which the IPSec peer is reachable.

- Following the creation of a policy and the binding of the proposal and ACL to the policy, the policy itself can be applied to the physical interface upon which traffic will be subjected to IPSec processing. In the example given an IPSec tunnel is to be established between interface Gigabit Ethernet 0/0/1 of RTA and Gigabit Ethernet 0/0/1 of RTB. Under the interface view, the ipsec policy command is used to apply the created policy to the interface.
- One interface can only use one IPSec policy group. To apply a new IPSec policy group to the interface, the previous policy group must be firstly removed. An IPSec policy group that establishes an SA through IKE negotiation can be applied to multiple interfaces, whereas an IPSec policy group that establishes an SA in manual mode can be applied only to one interface. When sending a packet from an interface, the AR2200 matches the packet with each IPSec policy in the IPSec policy group. If the packet matches an IPSec policy, the AR2200 encapsulates the packet according to the policy. If not, the packet is sent without encapsulation.



## IPSec Policy Verification

```
[RTA]display ipsec policy
=====
IPSec policy group: "P1"
Using interface: GigabitEthernet0/0/1
=====
Sequence number: 10
Security data flow: 3001
Tunnel local address: 20.1.1.1
Tunnel remote address: 20.1.1.2
Qos pre-classify: Disable
Proposal name:tran1
...
```

- Policy must associate with the policy of the peering interface.

- Using the `display ipsec policy [brief | name policy-name [ seq-number ] ]` command, a specific IPSec policy or all IPSec policies can be viewed. The parameters of the IPSec policy displayed include the policy name and the sequence number of the policy, typically used to distinguish the priority of policies in a group where a lower sequence number represents a higher priority. The policy bindings such as the name of the proposal and ACL that is used to filter the traffic in a policy is displayed along with the local and remote tunnel addresses for the IPSec policy.



## IPSec Policy Verification

```
...
Inbound ESP setting:
    ESP SPI: 12345 (0x3039)
    ESP string-key: huawei
    ESP encryption hex key:
    ESP authentication hex key:
Outbound ESP setting:
    ESP SPI: 54321 (0xd431)
    ESP string-key: huawei
    ESP encryption hex key:
    ESP authentication hex key:
...
...
```

- Policy Key strings must match for communication to establish.

- Specific parameter settings are also defined within the display ipsec policy command for SA in both the inbound and outbound directions. The SPI references the SA in each direction which can be used as part of the troubleshooting process to ensure that the SPI on the outbound local interface matches to that set for the inbound remote interface on the peer interface (that matches the tunnel remote address). The key string will also be defined for both the inbound and outbound SA to which the same local and remote configuration assessment can be performed.



## Summary

- What is meant by a Security Association (SA)?
- What are the three possible actions that may be applied to IPSec filtered traffic?

- A security association can be understood as a unidirectional connection or management construct that defines the services that will be used in order to negotiate the connection's establishment. The connection relies on matching SA in both directions in order to successfully establish secure bidirectional communication.
- All traffic that is destined to be forwarded via the outbound interface on which IPSec is established will firstly be subjected to filtering via a Security Policy Database (SPD). This is generally in the form of an ACL that will determine whether the traffic should be either dropped (Discard), forwarded normally (Bypass), or forwarded over a secure IPSec boundary (Protect).

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# Generic Routing Encapsulation

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- Limitations within IPSec VPN restrict the ability for routes to be carried between disparate site-to-site based networks, and allowing only for static route solutions. GRE provides a mechanism for the encapsulation of packets of one protocol into packets of another protocol. The application of GRE is as such implemented as a primary solution to the IPSec VPN limitations, for which knowledge of GRE is necessary to complement the existing knowledge of IPSec VPN.



## Objectives

- Upon completion of this section, you will be able to:
  - Explain how GRE can be applied to provide various solutions.
  - Describe the principle behavior of GRE.
  - Configure GRE over IPSec.



## GRE Application

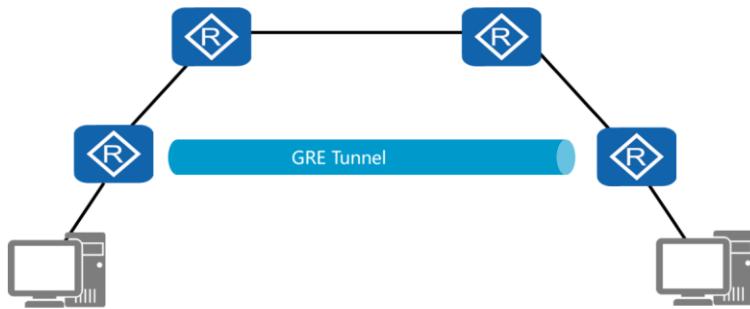


- Supports encapsulation of protocols over other protocols.
- Enables routing between remote and disparate networks.

- Cost effective solutions for private internetwork communication have utilized technologies such as IPSec to enable shared public networks to act as a medium for bridging the gap between remote offices. This solution provides secure transmission to IP packet payloads but does not enable routing protocols such as RIP and OSPF to be carried between the two tunnel endpoints. GRE was originally developed as a means of supporting the transmission of protocols such as IPX through IP networks, however over time the primary role of GRE has transitioned towards the support of routing based protocol tunneling.



## GRE Scaling Solution for IGP



- Application allows for increased scalability of IGP networks.
- Capable of building a tunnel to resolve hop count limitations.

- GRE can be applied within an autonomous system to overcome certain limitations found in IGP protocols. RIP routes offer for example a hop limit of 15 hops beyond which all routes are subjected to the count to infinity rule. The application of GRE in such environments will allow for a tunnel to be created over which routing updates can be carried, allowing the scale of such distance vector routing to be increased significantly where necessary.



## IPSec VPN support for GRE

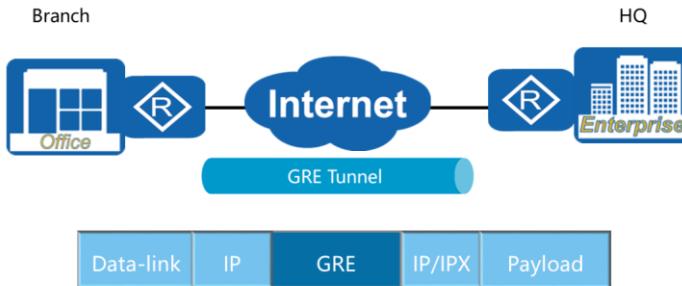


- GRE contains no means for confidentiality of GRE payloads.
- IPSec can be employed to provide confidentiality to GRE.

- One of the major limitations of GRE is in its lack of ability to secure packets as they are carried across a public network. The traffic carried within a GRE tunnel is not subjected to encryption since such features are not supported by GRE. As a result IPSec solutions are employed together with GRE to enable GRE tunnels to be established within IPSec tunnels to extend features including integrity and confidentiality to GRE.



## GRE Packet Encapsulation & Decapsulation

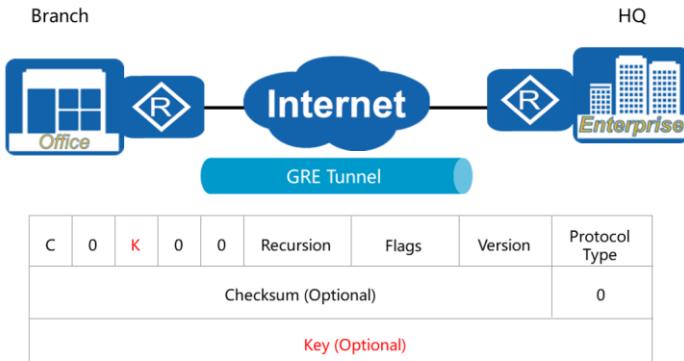


- A GRE header is inserted into the packet to build a tunnel.
- A virtual network is built over the physical network.

- After receiving a packet from the interface that is connected to the private network, the packet is delivered to the private network protocol module for processing. The private network protocol module checks the destination address field in the private network packet header, searches the outgoing interface in the routing table or forwarding table of the private network, and determines how to route this packet. If the outgoing interface is the tunnel interface, the private network protocol module sends the packet to the tunnel module.
- After receiving the packet, the tunnel module firstly encapsulates the packet according to the protocol type of the passenger protocol and checksum parameters are configured for the current GRE tunnel. That is, the tunnel module adds a GRE header to the packet. It next adds an IP header according to the configuration. The source address of this IP header is the source address of the tunnel; the destination address of the IP header is the destination address of the tunnel, following which the packet is delivered to the IP module. The IP module searches for the appropriate outgoing interface in the public network routing table based on the destination address in the IP header, and forwards the packet. The encapsulated packet is then transmitted over the public network.
- After receiving the packet from the interface that is connected to the public network, the egress interface analyzes the IP header, determines the destination of the packet and discovers the Protocol Type field is 47, indicating that the protocol is GRE.
- The egress interface delivers the packet to the GRE module for processing. The GRE module removes the IP header and the GRE header, and learns from the Protocol Type field in the GRE header that the passenger protocol is the protocol run on the private network, the GRE module then delivers the packet to this protocol. The protocol itself handles the packet as with any ordinary packet.



## GRE Key Authentication

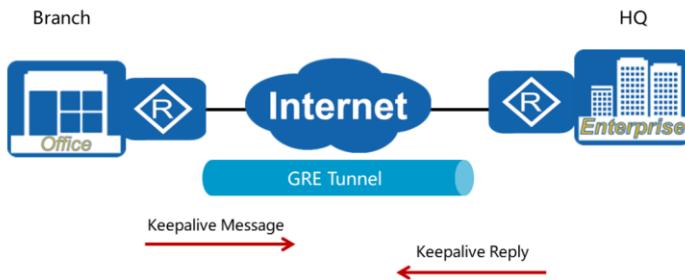


- Key field in GRE provides a means of optional authentication.

- Key authentication indicates the authentication on a tunnel interface. This security mechanism can prevent the tunnel interface from incorrectly identifying and receiving the packets from other devices. If the K bit in the GRE header is set to 1, the Key field is inserted into the GRE header. Both the receiver and the sender perform the key authentication on the tunnel.
- The Key field contains a four-byte value, which is inserted into the GRE header during packet encapsulation. The Key field is used to identify the traffic in the tunnel. Packets of the same traffic flow have the same Key field. When packets are decapsulated, the tunnel end identifies the packets of the same traffic according to the Key field. The authentication can be passed only when the Key fields set on both ends of the tunnel are consistent, otherwise the packets are discarded.



## GRE Keepalive

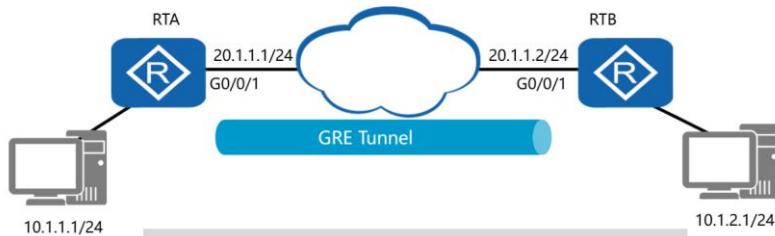


- Allows the status of a GRE tunnel to be monitored for changes.
- Keepalives that are not replied result in GRE tunnel tear down.

- The keepalive detection function is used to detect whether the tunnel link is in the keepalive state at any time, that is, whether the peer of the tunnel is reachable. If the peer is not reachable, the tunnel is disconnected to prevent black holes occurring. After the Keepalive function is enabled, the local end of the GRE tunnel periodically sends a keepalive detection packet to the peer. If the peer is reachable, the local end receives a reply packet from the peer. The keepalive function will operate as long as at least one end is configured with the keepalive function, the peer does not need to have the keepalive function enabled. If the peer receives a keepalive detection packet, it sends a reply packet to the local end, irrespective of whether it is configured with the keepalive function.
- After the keepalive function is enabled, the source of a GRE tunnel creates a counter, periodically sends the keepalive detection packets, and counts the number of detection packets. The number increments by one after each detection packet is sent.
- If the source receives a reply packet before the counter value reaches the preset value, the source considers that the peer is reachable. If the source does not receive a reply packet before the counter reaches the preset value that defines the number of retries, the source considers that the peer is unreachable and will proceed to close the tunnel connection.



## GRE Configuration



```
[RTA]interface Tunnel 0/0/1
[RTA-Tunnel0/0/1]ip address 40.1.1.1 24
[RTA-Tunnel0/0/1]tunnel-protocol gre
[RTA-Tunnel0/0/1]source 20.1.1.1
[RTA-Tunnel0/0/1]destination 20.1.1.2
[RTA-Tunnel0/0/1]quit
[RTA]ip route-static 10.1.2.0 24 Tunnel 0/0/1
```

- The establishment of GRE requires initially the creation of a tunnel interface. After creating a tunnel interface, specify GRE as the encapsulation type, set the tunnel source address or source interface, and set the tunnel destination address. In addition, set the tunnel interface network address so that the tunnel can support routes. Routes for a tunnel must be available on both the source and destination devices so that packets encapsulated with GRE can be forwarded correctly. A route passing through tunnel interfaces can be a static route or a dynamic route.
- One other key point that should be taken into account during the configuration phase is the MTU. GRE results in an additional 24 bytes of overhead being applied to the existing headers and payload, resulting in the potential for unnecessary fragmentation to occur. This can be resolved by adjusting the MTU size using the `mtu <mtu>` command to allow packets to compensate for the additional overhead created. An MTU of 1476 is considered a sufficient adjustment to the MTU, to reduce additional load on the interface.



## Configuration Validation

```
[RTA]display interface Tunnel 0/0/1
Tunnel0/0/1 current state : UP
Line protocol current state : UP
Last line protocol up time : 2016-03-21 13:37:38
Description:HUAWEI, AR Series, Tunnel0/0/1 Interface
Route Port, The Maximum Transmit Unit is 1476
Internet Address is 40.1.1.1/24
Encapsulation is TUNNEL, loopback not set
Tunnel source 20.1.1.1 (GigabitEthernet0/0/1), destination 20.1.1.2
Tunnel protocol/transport GRE/IP, key disabled
keepalive disabled
Checksumming of packets disabled
....
```

- Enables confirmation of tunnel status and parameters.

- The running status and routing information for the GRE tunnel interface can be observed following the creation of the tunnel interface. An active tunnel interface can be confirmed when the tunnel's current state and link layer (line) protocol state are both showing as UP. Any adjustment to the existing MTU can be identified to ensure that additional overhead is not created as a result of fragmentation. The Internet address represents the configured tunnel IP address at the egress interface. The tunnel source and destination displays the IP addresses of the physical interface that are being used, over which the GRE tunnel is established.



## GRE Routing Table Validation

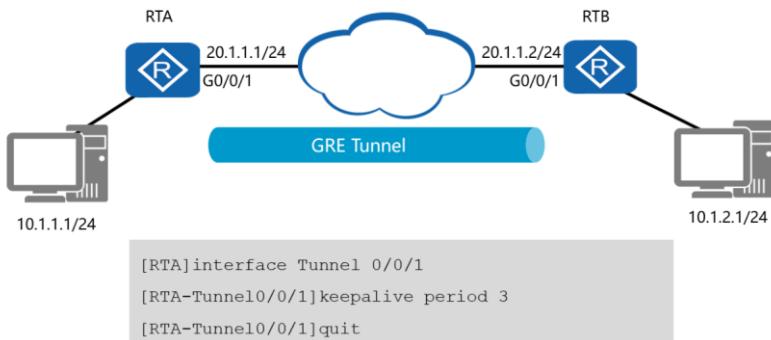
```
[RTA]display ip routing-table
Route Flags: R - relay, D - download to fib
-----
Routing Tables: Public Destinations : 13 Routes : 14
Destination/Mask Proto Pre Cost Flags NextHop Interface
.....
10.1.2.0/24 Static 60 0 RD 40.1.1.2 Tunnel 0/0/1
.....
```

- An entry in the routing table verifies the tunnel establishment.
- Routes for GRE can be static or dynamic.

- The reachability of networks over the GRE tunnel can be determined based on the IP routing table that is viewed using the display ip routing table command. In this case it is possible to identify the internal source network address of a local network, and its ability to reach the destination network address of a remote network over a GRE based tunnel interface. The destination address refers to the network address that is reached over the GRE tunnel, while the next hop address references the IP address of the GRE remote tunnel interface.



## Enabling the Keepalive Function



- Keepalives can define message interval and number of retries.
- Function only required to be configured on one tunnel interface.

- The keepalive function can be enabled on the GRE tunnel interface by specifying the keepalive [period <period> [retry-times <retry-times> ] ] command, where the period determines the number of seconds taken before each keepalive is sent, this by default is set to a period of 5 seconds. The number of times for which a keepalive will be sent is defined using the retry-times <retry-times> parameter, which by default is set to 3 tries. If the peer fails to respond within the number of retries defined, communication between the two ends of the tunnel will be considered to have failed and the GRE tunnel will proceed to be torn down.



## Configuration Validation

```
[RTA]display interface Tunnel 0/0/1
Tunnel0/0/1 current state : UP
Line protocol current state : DOWN
Description:HUAWEI, AR Series, Tunnel0/0/1 Interface
Route Port, The Maximum Transmit Unit is 1476
Internet Address is 40.1.1.1/24
Encapsulation is TUNNEL, loopback not set
Tunnel source 20.1.1.1 (GigabitEthernet0/0/1), destination 20.1.1.2
Tunnel protocol/transport GRE/IP, key disabled
keepalive enable period 3 retry-times 3
Checksumming of packets disabled
.....
```

- Keepalive enablement can be verified from the tunnel interface.

- The display interface tunnel command can again be applied to assess whether the tunnel is still active following a loss of communication between the GRE tunnel endpoints. In the example it is found that the link layer status of the GRE tunnel has transitioned to a down state denoting that a failure in communication has occurred. The keepalive function is currently enabled and displays a keepalive period of three seconds between each message sent. The number of retries has also been determined as allowing a loss of replies up to three times.



## Summary

- What is the primary application for using GRE?
- What is the difference between the Internet Address and the Tunnel source in the display interface tunnel command?

- GRE provides a means for establishing dynamic routing between remote networks that commonly belong to a single administrative domain such as in the case of branch offices. IPSec VPN is generally used to provide a site-to-site private tunnel over which routing dynamic information may wish to be transmitted, however IPSec VPN tunnels will not support the forwarding of routing updates over IPSec directly, therefore GRE is implemented.
- The Internet address represents a virtual tunnel network address over which the GRE is established, while the tunnel source references the entry-point at which the tunnel is established.

A blue-tinted silhouette of several business people standing in a modern office environment with large windows and a glass floor.

Thank You

[www.huawei.com](http://www.huawei.com)



# Simple Network Management Protocol

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- Management framework solutions for TCP/IP networks were introduced as hardware and software increased, in order to support rapid network growth. SNMP was originally adapted from a simpler SGMP protocol for use as the basis for common network management throughout the system. SNMP has since experienced version revisions, however remains the standard protocol for network management. The SNMP framework, as well as the supporting Management Information Base act as the foundation for network management, and are introduced in support of a well-rounded understanding of the network management framework for TCP/IP.



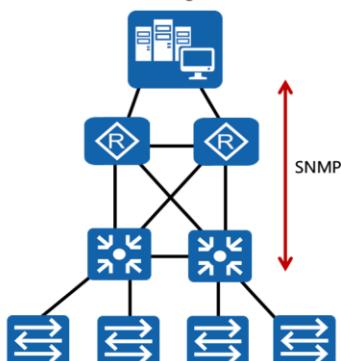
## Objectives

- Upon completion of this section, you will be able to:
  - Describe the SNMP architecture and messaging behavior.
  - Describe the function of the Management Information Base (MIB).
  - Configure general SNMP parameters and traps.



## SNMP Application

Network Management Station

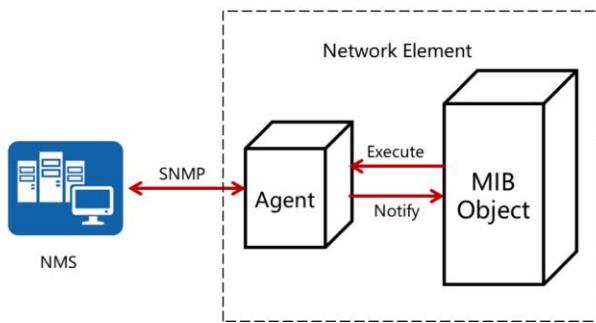


- SNMP is used to communicate management information between network management stations and network elements.

- The Simple Network Management Protocol (SNMP) is a network management protocol widely used in the TCP/IP network. SNMP is a method of managing network elements using a network console workstation which runs network management software.
- SNMP may be used to achieve a number of communicative operations. The Network Management Station (NMS) relies on SNMP to define sources for network information and obtain network resource information. SNMP is also used to relay reports in the form of trap messages to the NMS so that the station can obtain network status in near real time, to allow the network administrator to quickly take action in the event of system discrepancies and failures.
- SNMP is largely used to manage application programs, user accounts, and write/read permissions (licenses) etc., as well as to manage the hardware that makes up the network, including workstations, servers, network cards, routing devices, and switches. Commonly, these devices are located far from the central office where the network administrator is based. When faults occur on the devices, it is expected that the network administrator can be notified automatically of the faults. SNMP effectively operates as a communications medium between the network elements and the network administrator/NMS.



## SNMP Architecture

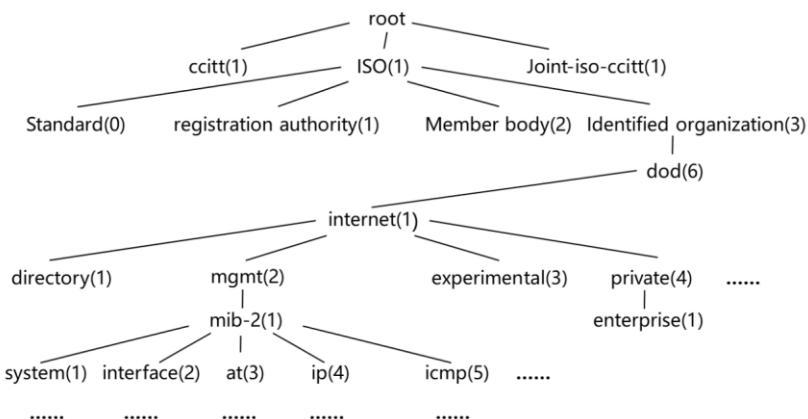


- Agents exist within network elements through which NMS interact to retrieve or alter parameter variables in the MIB.

- Network elements such as hosts, gateways, terminal servers etc., contain two important components that support the network management functions requested by the network management stations. The management agent resides on the network element in order to retrieve (get) or alter (set) variables.
- Network Management Stations (NMS) associate with management agents that are responsible for performing the network management functions requested by the NMS. The MIB stores a number of variables associated with the network element, with each of these variables being considered an MIB object. The exchange of SNMP messages within IP requires only the support of UDP as an unreliable datagram service for which each message is independently represented by a single transport datagram.



## MIB Objects

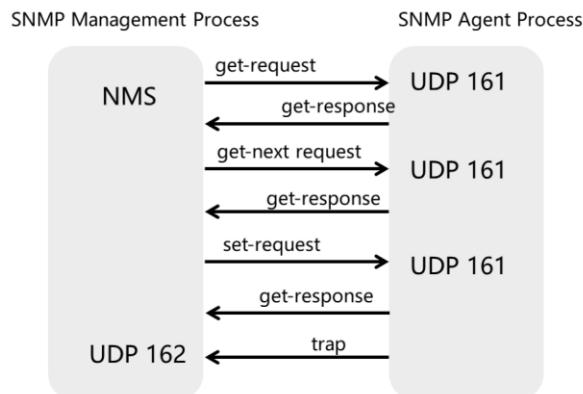


- The MIB acts as a virtual store for the management of objects.

- A Management Information Base (MIB) specifies the variables maintained by network elements. These variables are the information that can be queried and set by the management process. A MIB presents a data structure, collecting all possible managed objects over the network. The SNMP MIB adopts a tree structure similar to that found in a Domain Name System (DNS).
- The object naming tree has three top objects: ISO, ITU-T (originally CCITT), and the joint organizations branch. Under the ISO, there are four objects among which number 3 is the identified organization. A sub-tree of the US Department of Defense dod (6) is defined under the identified organization (3) under which the Internet (1) sub tree is located. The object under the Internet is mgmt (2). What follows mgmt (2) is MIB-II, originally MIB until 1991 when the new edition MIB-II was defined. The tree path itself can be defined as an object identifier (OID) value {1.3.6.1.2.1}.



## SNMP Operation

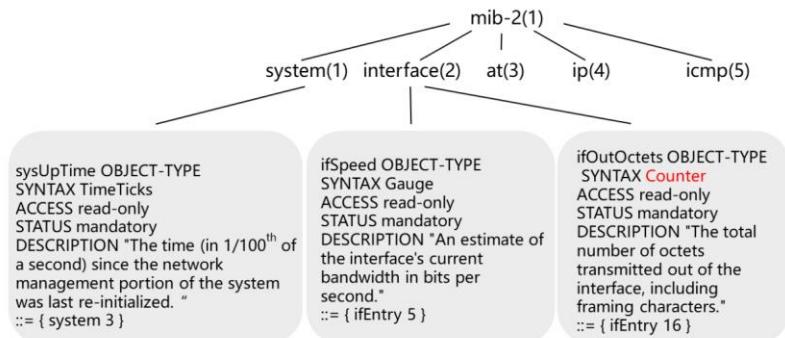


- Requests are received by an SNMP agent on UDP port 161.

- SNMP defines five types of Protocol Data Units (PDUs), namely, SNMP packets, to be exchanged between the management process and the agent process. The get-request operation indicates that the management process reads one or more parameter values from the MIB of the agent process. The get-next-request indicates that the management process reads the next parameter value in the lexicographic order from the MIB of the agent process. The set-request indicates that the management process sets one or more parameter values in the MIB of the agent process. The get-response returns one or more parameter values. This operation is performed by the agent process. It is the response to the preceding three operations. Lastly is the trap function which is actively sent by the agent process to inform the management process of important or critical events.



## SNMPv2c

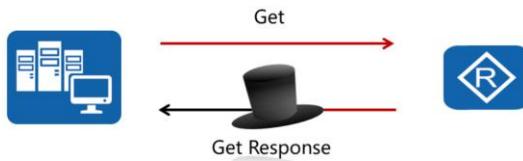


- New Get-bulk request PDU and inform request included.
- 64 bit counters introduced to prevent counter wrap.

- SNMPv1 is the original application protocol by which the variables of an agent's MIB may be inspected or altered. The evolution of SNMP involved not only changes to the protocol but also the MIB that was used. New objects were defined in the MIB resulting in MIB-II (or MIB-2) being defined, including for example `sysContact`, `sysName`, `sysLocation`, `sysServices` to provide contact, administrative, location, and service information regarding the managed node in the system group, and `ipRouteMask`, `ipRouteMetric5`, and `ipRouteInfo` objects included as part of the IP route table object.
- The transition to SNMP version 2 involved a number of revisions that resulted in SNMPv2c being developed including the introduction of a new PDU type in the form of `GetBulkRequest-PDU` to allow information from multiple objects to be retrieved in a single request and the `Inform Request`, a manager to manager communication PDU, used where one manager sends information from an MIB view to another manager. Specific objects also use counters as a syntax which in SNMP version 1 represented a 32 bit value. This meant that in given objects such as the byte count of interfaces it was easy for the counter to complete a full cycle of the values and wrap, similar to the odometer that measures mileage in vehicles.
- Using the 32 bit counter, octets on an Ethernet interface transmitting at 10Mbps would wrap in 57 minutes, at 100Mbps the counter would wrap in 5.7 minutes, and at 1Gbps it would take only 34 seconds before the counter fully cycled. Objects are commonly polled (inspected) every 1 or 5 minutes, and problems arise when counters wrap more than once between object polling as a true measurement cannot be determined.
- To resolve this, new counters were defined in SNMP version 2c in the form of 64 bit counters for any situations where 32 bit counters wrap too fast, which translated to any interface that counts faster than 650 million bits per second. In comparison, using a 64 bit counter for counting octets on a 1Tbps (1,000 Gbps) will wrap in just under 5 years, and it would take an 81,000,000 Tbps link to cause a 64-bit counter to wrap in 30 minutes.



## SNMPv3

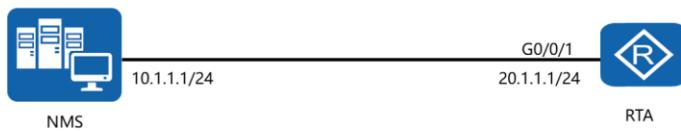


- SNMPv3 security mechanisms support data integrity, data origin authentication, confidentiality and timely message delivery.

- One of the key improvements to SNMPv3 is with regards to security of the transmission of MIB object information. Various threats can be identified. These include modification of object information from an unauthorized entity during transit, the performing of unauthorized management operations by users masquerading as another authorized user; eavesdropping on message exchanges and the modification of the message stream through means such as message replay.
- SNMP enhances security through applying four principle measures. Data integrity is applied to ensure that data has not been altered or destroyed in an unauthorized manner, nor have data sequences been altered to an extent greater than can occur non-maliciously.
- Data origin authentication is supported to ensure that the claimed identity of the user on whose behalf received data was originated is corroborated using MD5 and SHA-1. Data confidentiality is applied to ensure information is not made available or disclosed to unauthorized individuals, entities, or processes. Additionally, solutions for limited replay protection provide a means of ensuring that a message, whose generation time is outside of a specified time window, is not accepted.



## SNMP Configuration



```
[RTA] snmp-agent
[RTA] snmp-agent sys-info version v2c
[RTA] snmp-agent trap enable
Info: All switches of SNMP trap/notification will be open. Continue?
[Y/N] :Y
[RTA] snmp-agent trap source GigabitEthernet 0/0/1
```

- The SNMP agent is an agent process on a device on the network. The SNMP agent maintains managed network devices by responding to NMS requests and reporting management data to the NMS. To configure SNMP on a device, the SNMP agent must be enabled, for which the `snmp-agent` command is applied.
- The `snmp-agent sys-info` command sets the SNMP system information and is also used to specify the version(s) of SNMP that are supported, where `snmp-agent sys-info version [ [ v1 | v2c | v3 ] * | all ]` is used to achieve this, and should be noted that all versions of SNMP are supported by default. The `snmp-agent trap enable` command, activates the function of sending traps to the NMS, following which the device will proceed to report any configured events to the NMS.
- In addition it is necessary to specify the interface via which trap notifications will be sent. This should be the interface pointing towards the location of the NMS, as in the example where the NMS is reached via interface Gigabit Ethernet 0/0/1.



## Configuration Validation

```
[RTA]display snmp-agent sys-info
The contact person for this managed node:
R&D Shenzhen, Huawei Technologies Co., Ltd.

The physical location of this node:
Shenzhen China

SNMP version running in the system:
SNMPv2c
```

- Using the `display snmp-agent sys-info` command displays contact information of personnel responsible for the system maintenance, the physical location of the device, and currently supported SNMP version(s). The given information in the example represents the typical default system information found within Huawei AR2200 series routers, however this can be altered through the use of the `snmp-agent sys-info [contact | location | version ]` parameters to reflect contact and location details relevant to each individual device.



## Summary

- Which version(s) of SNMP is/are enabled by default?
- What is the destination port number that is used by an agent to forward traps to a Network Management Station?

- In the Huawei AR2200 series router, all versions of SNMP (SNMPv1, SNMPv2c and SNMPv3) are enabled by default.
- The agent forwards trap messages to the Network Management Station (NMS) using UDP destination port 162.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# Introducing IPv6 Networks

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- With the gradual exhaustion of the IPv4 address space, new solutions for continued address space were needed. Temporary measures in the form of NAT were applied, however long term solutions were required. The IPv6 addressing architecture is a developing solution for IP to provide for the next generation of networks and beyond. The transition to an all IPv6 architecture, while progressive, requires a major overhaul of many protocols and applications as well as standards. The IPv6 network however aims to resolve many limitations within the current TCP/IP suite, most notably addressing the need for integrated security measures and the streamlining of protocols to minimize overhead. A deep knowledge of the IPv6 architecture is required by engineers as IPv6 continues to evolve as an integral part of the enterprise network.



## Objectives

- Upon completion of this section, you will be able to:
  - Explain the characteristics of IPv6.
  - Explain the IPv6 address format and addressing types.
  - Describe the process for IPv6 stateless address auto-configuration.



## IPv6 Addressing

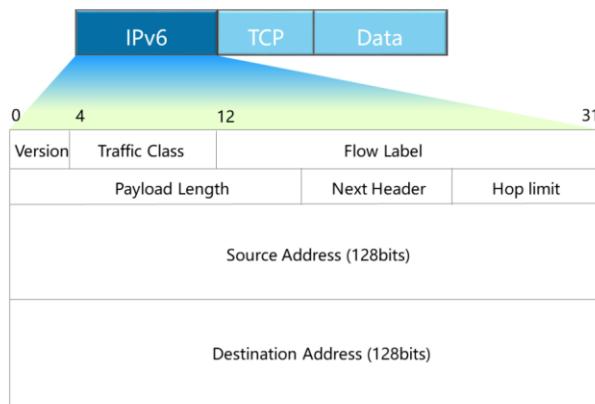
Version	Address size	Total Number of Addresses
IPv4	32 bit	4,294,967,296
IPv6	128 bit	340,282,366,920,938,463,463,374,607,431,768,211,456

- Exhaustion of the limited IPv4 address space.
- IPv6 addressing implemented to resolve address shortages.

- As a set of specifications defined by the Internet Engineering Task Force (IETF), Internet Protocol version 6 (IPv6) is the next-generation network layer protocol standard and the successor to Internet Protocol version 4 (IPv4). The most obvious difference between IPv6 and IPv4 is that IP addresses are lengthened from 32 bits to 128 bits. In doing so, IPv6 is capable of supporting an increased number of levels of addressing hierarchy, a much greater number of addressable nodes, and simpler auto-configuration of addresses.
- The existing IPv4 address range was implemented at a time when such networks as ARPANET and the National Science Foundation Network (NSFNET) represented the mainstream backbone network, and at a time when IPv4 was considered more than ample to support the range of hosts that would connect to these forming networks. The unforeseen evolution of the Internet from these ancestral networks resulted in the rapid consumption of the IPv4 address space of 4.3 billion addresses (of which many are reserved), for which counter measures in the form of NAT and CIDR were put in place to alleviate the expansion, and give time for a more permanent solution to be formed. Additionally, early IPv4 network address allocation was highly discontiguous making it difficult for addressing to be clustered into effective address groups and ease the burden on global IP routing tables used in autonomous system based routing protocols such as BGP.
- Eventually however the IP network is expected to shed its IPv4 addressing to make way for IPv6 and provide the addressing capacity of over 340 undecillion unique addresses, considered more than necessary for continued IP network growth. Along with this, address allocation by the Internet Assigned Numbers Authority (IANA) ensures that address allocation for IPv6 is contiguous for efficient future management of IP routing tables.



## IPv6 Header Format

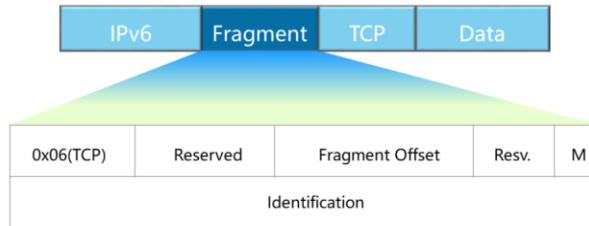


- IPv6 header has been streamlined to reduce overhead.

- The IPv6 header required some level of expansion to accommodate the increased size of the IP address, however many other fields that were considered redundant in many cases of packet forwarding were removed in order to simplify the overall design of the IPv6 header, and optimize the level of overhead that is necessary during each packet that is forwarded across the network. In comparison with the IPv4 packet header, the IPv6 packet header no longer contains an Internet Header Length (IHL), identifier, flags, fragment offset, header checksum, options, or padding fields, but instead carries a flow label field.
- The term flow can be understood as a continuous stream of unicast, multicast or anycast packets relating to a specific application or process that originate from a given source and are transmitted to a given destination or multiple destinations. The natural behavior of a packet switched network means that such traffic flows may be carried over multiple paths and arrive at the intended destination out of sequence. System resources are then required to re-sequence the packet flow before it can be processed by the upper layers.
- The flow label intends to identify such traffic flows, together with the source and destination address fields, to maintain a common path across networks for all packets associated with the traffic flow. In doing so packets are capable of maintaining sequence over the IP network, and this optimizes process efficiency.
- As IP network convergence introduces technologies such as voice that relies on a circuit switched traffic flow behavior and near real time performance, the importance of the flow label field becomes more paramount.
- Various options can also now be supported without changing the existing packet format, with the inclusion of extension header information fields. These extension header fields are referenced through the Next Header field that replaces the protocol field found in IPv4, in order to provide additional flexibility in IPv6.



## IPv6 Extension Header

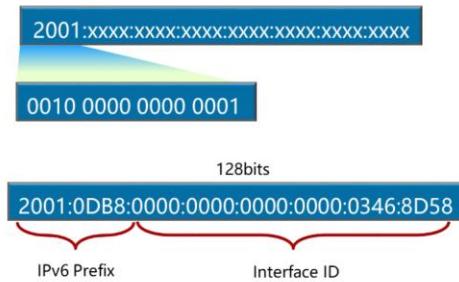


- Extension headers are used to support parameters that are not required in every IP packet, such as fragmentation and IPSec.

- The extension header is one of the primary changes to IPv6 for enabling the IPv6 header to be more streamlined. There are a number of extension headers that exist in IPv6 and are referenced directly through a hexadecimal value in the IPv6 Next Header field, in the same way that ICMP, TCP, UDP headers are referenced in the protocol field of the IPv4 header. Each extension header also contains a next header field to reference the next header following the processing of the extension header. Multiple extension headers can be associated with a packet and each referenced in sequence. The list of extension headers and the sequence in which the extension headers are processed are as follows.
  - IPv6 header
  - Hop-by-Hop Options header
  - Destination Options header
  - Routing header
  - Fragment header
  - Authentication header
  - Encapsulating Security Payload header
  - Destination Options header
  - Upper-layer header (as in IPv6 tunneling)
  - Each extension header should occur at most once, except for the Destination Options header which should occur at most twice (once before a Routing header and once before the upper-layer header). Another key aspect of the extension headers is the introduction of IPSec protocols in the form of AH and ESP to enhance the overall security of IPv6.



# IPv6 Address Architecture

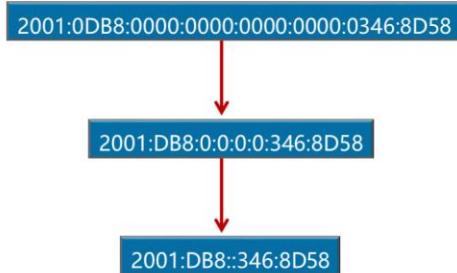


- IPv6 address consists of a prefix and an interface identifier.
- Addresses are commonly displayed in hexadecimal format.

- The IPv6 address is a 128 bit identifier that is associated with an interface or a set of interfaces as (in the case of anycast address assignment covered later). The address notation due to size is commonly defined in hexadecimal format and written in the form of xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx, where the xxxx relates to the four hexadecimal digits that translate to a 16 bit binary value. The set of 8\*16 bit grouped hexadecimal values comprises the 128 bit IPv6 address.
- The address itself is made up of an IPv6 prefix that is used to determine the IPv6 network, followed by an interface ID. The basic addressing scheme represents a two layer architecture of prefix and interface ID, however IPv6 address types may contain many sublevels that enable an extensive hierarchical addressing architecture to be formed in the IPv6 network, to support effective grouping and management of users and domains.



## IPv6 Address Condensing



- Addresses can be condensed by removing the leading zeroes.
- The :: operator will further condense strings of zero values.

- The extensive 128 bit size of the IPv6 address means that the general written notation can be impractical in many cases. Another factor that will remain common for quite some time is the presence of long strings of zero bits, primarily due to the unfathomable enormity of the address scale involved. This however brings a means for simplification and practicality when writing IPv6 addresses that contain zero bits. One initial means of simplification of the address scope is through the removal of any leading zero values that are found in each set of 16 bit hexadecimal values.
- A special syntax is available to further compress the zeros and is represented in the form of a double colon or "::". This value indicates one or more 16 bit groups of zeros. In order to enable the string size to be determinable however, the "::" can only appear once in an address. The double colon "::" can be used to compress leading zeros in an address as displayed in the given example.



## IPv6 Address Reservations

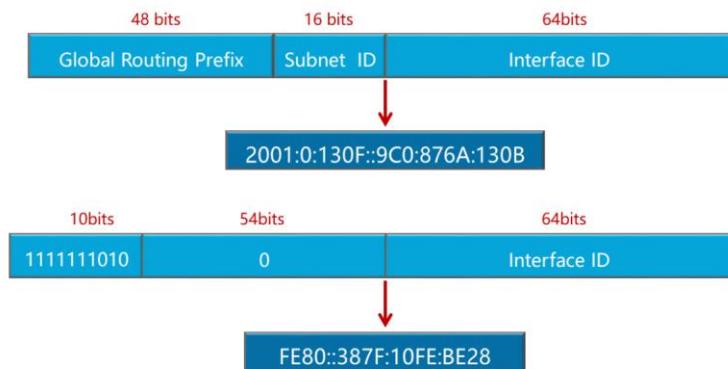
Address Range	Description
2000::/3	Current Global Unicast Range
2001:0DB8::/32	Reserved for Documentation
FE80::/10	Link Local Unicast Address Range
FF00::/8	Multicast Address Range
::/128	Unspecified Address
::1/128	Loopback Address

- Address ranges have been allocated in IPv6 for unicast and multicast, along with special addresses for operational support.

- The IPv6 address space has yet to be completely defined firstly due to the magnitude of the addressing space, for which it is not likely to be required anytime in the near future, secondly the formulation of addressing schemes is still somewhat in an infancy stage with much discussion occurring with regards to the address types that should be applied.
- Currently a fraction of the address range has been sectioned for use in the form of a 2000::/3 address range that represents a global unicast address range. This can be understood to be any address that is routable in the public IP network. The governing body IANA (now a part of ICANN) is responsible for distributing portions of this address range to various Regional Internet Registries (RIR) that manage addressing for one of five regions in the world. Supersets of IPv6 addresses in the form of 2400::/12 (APNIC), 2600::/12 (ARIN), 2800::/12 (LACNIC), 2A00::/12 (RIPE NCC) and 2C00::/12 (AfriNIC) have been allocated, allowing the potential for addressing for a given region to be referenced and managed in the meantime using a single address prefix. Also contained within the 2000::/3 prefix range are reserved address fields, including 2001:0DB8::/32 that is used to reference fictional addressing within documentation.
- Multicast addressing is defined within IPv6 as FF00::/8, with much of the addressing scope being reserved to specific address ranges (such as link local) and in support of routing protocols, in a manner similar to how multicast addresses are used in IPv4. One major change to IPv6 is that there are no broadcast addresses, their function being superseded by multicast addresses, which reduces unnecessary processing by end stations at the MAC layer within IPv6 networks, since hosts are able to filter multicast MAC address groups.
- Some special addresses also exist within IPv6 including the unspecified address ::/128 which represents an interface for which there is no IP address currently assigned. This should not be confused however with the default IP address ::/0 which is used as a default address value for any network in the same way the 0.0.0.0/0 default address is used within IPv4. For the loopback address (127.0.0.1), this is defined in IPv6 as the reserved address ::1/128.



## IPv6 Addressing – Unicast



- Global unicast address prefixes are used for public networks.
- Prefix ranges are reserved for various IP transmission methods.



## IPv6 Addressing – Multicast



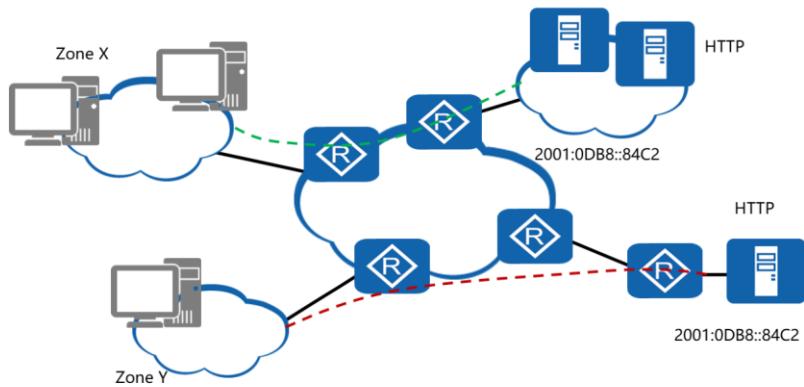
Address Range	Description
FF02::1	All Nodes Addresses (Link Local)
FF02::2	All Routers Addresses (Link Local)

- Multicast addresses are distinguished by an FF00::/8 prefix.
- Select multicast address groups are reserved for protocol use.

- A multicast address is an identifier for a set of interfaces (typically belonging to different nodes). A packet sent to a multicast address is delivered to all interfaces identified by that address. The multicast address range is determined from a FF00::/8 address prefix and is clearly distinguishable from the first 8 bits which are always set to 11111111. The address architecture for multicast addresses comprises of flags, scope and group ID fields.
- The flags field though 4 bits, is currently used to support identification of whether a multicast address is well known (as in assigned and recognized by the global internet numbering authority) or transient, meaning the multicast address is temporarily assigned. A second bit value is used within the flags field to determine whether or not the multicast address is based on the network prefix. The remaining higher order two bits are reserved and remain set to 0.
- The scope defines a number of potential ranges to which the multicast is applied. Generally this may refer to a node local scope FF01::, or to the link local scope FF02:: which refers to the listeners within the scope of the link-layer, such as an Ethernet segment for which a gateway defines the boundary. Common well known link local multicast addresses include FF02::1 which is used to forward to all nodes, and FF02::2 that allows multicast transmission to all router addresses.



## IPv6 Addressing – Anycast

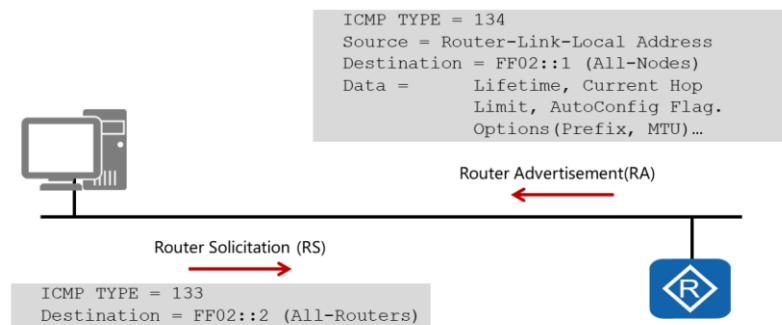


- Anycast allows multiple instances of a service to be associated with a single address, enabling a variety of service applications.

- Anycast represents an identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the "nearest" one, according to the routing protocols' measure of distance). The anycast architecture commonly involves an anycast initiator and one or multiple anycast responders.
- The anycast initiator may commonly be a host that is enquiring to a service such as in the case of a DNS lookup, or requesting the retrieval of specific data, such as HTTP web page information. Anycast addresses are in no way distinguishable from unicast addresses, with the exception that multiple instances of the same address are applied to a given device.
- The application for anycast over traditional methods of addressing brings many potential benefits to enterprise network operation and performance. Service redundancy is one such application for anycast that allows a service such as HTTP to be used across multiple servers, for which the same address is associated with one or more servers that operate as anycast responders. In the event that service failure occurs on one server, clients would need traditionally to search for the address of another server and then restart/reestablish communication.
- Using anycast addressing, the anycast initiator automatically establishes communication with another server that is using the same address, providing an effective redundancy solution.
- In terms of performance, multi-server access can be applied, where an anycast initiator establishes a connection to the same unicast address, each connection automatically establishes to multiple servers that specify the same anycast address such as in the case of a client browsing a company web page.
- The client will download the html file and individual image files from separate mirror servers. As a result, the anycast initiator utilizes bandwidth from multiple servers to download the web page much more quickly than is possible with a unicast approach.



## IPv6 Stateless Address Auto-configuration



- Hosts are capable of generating IPv6 addresses independently.
- Router Advertisements deliver network parameter information.

- Upon physically establishing with an IPv6 network, hosts must establish a unique IPv6 address and associated parameters such as the prefix of the network. Routers send out Router Advertisement messages periodically, and also in response to Router Solicitations (RS) to support router discovery, used to locate a neighboring router and learn the address prefix and configuration parameters for address auto-configuration.
- IPv6 supports stateless address auto-configuration (SLAAC) that allows hosts to obtain IPv6 prefixes and automatically generate interface IDs without requiring an external service such as DHCP. Router Discovery is the basis for IPv6 address auto-configuration and is implemented through two message formats.
- Router Advertisement (RA) messages allow each router to periodically send multicast RA messages that contain network configuration parameters, in order to declare the router's existence to Layer 2 hosts and other routers. An RA message is identified by a value of 134 in the type field of the message. Router Solicitation (RS) messages are generated after a host is connected to the network.
- Routers will periodically send out RA messages however should a host wish to prompt for an RA message, the host will send an RS message. Routers on the network will generate an RA message to all nodes to notify the host of the default router on the segment and related configuration parameters. The RS message generated by a host can be distinguished by the type field which contains a value of 133.

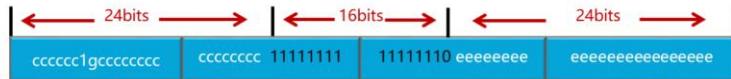


## EUI-64 for IP Stateless Address Auto-configuration

48-bit MAC address



EUI-64 generated interface ID

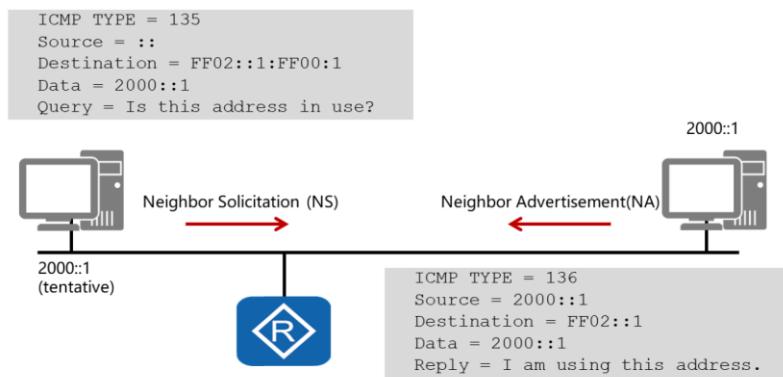


- A host MAC address is injected with 16 bit ‘FF’ ‘FE’ values, as well as the 7th bit is negated, to generate a 64-bit interface identifier for the IPv6 address.

- In order to facilitate communication over an IPv6 network, each interface must be assigned a valid IPv6 address. The interface ID of the IPv6 address can be defined either through manual configuration by a network administrator, generated through the system software, or generated using the IEEE 64-bit Extended Unique Identifier (EUI-64) format.
- Due to practicality, the most common means for IPv6 addressing is to generate the interface ID using the EUI-64 format. IEEE EUI-64 standards use the interface MAC address to generate an IPv6 interface ID. The MAC address however represents a 48-bit address whilst the required interface ID must be composed of a 64-bit value. The first 24 bits (expressed by c) of the MAC address represent the vendor (company) ID, while the remaining 24 bits (expressed by e) represents the unique extension identifier assigned by the manufacturer.
- The higher seventh bit in the address represents a universal/local bit to enable interface identifiers with universal scope. Where this value is 0, the MAC address is unique globally. During conversion, the EUI-64 process inserts two octet “0xFF” and “0xFE” values totaling 16 bits between the vendor identifier and extension identifier of the MAC address, and the universal/local bit 0 is changed to 1 to indicate that the interface ID now represents a globally unique address value. A 64-bit interface ID is created and associated with the interface prefix to generate the IPv6 interface address.



## IPv6 Stateless Address Auto-configuration DAD



- Duplicate Address Detection (DAD) is used in IPv6 to verify that an address is unique before it is applied to the host interface.

- Before an IPv6 unicast address is assigned to an interface, Duplicate Address Detection (DAD) is performed to check whether the address is used by another node. DAD is required if IP addresses are configured automatically. An IPv6 unicast address that is assigned to an interface but has not been verified by DAD is called a tentative address. An interface cannot use the tentative address for unicast communication but will join two multicast groups: ALL-nodes multicast group and Solicited-node multicast group.
- A Solicited-Node multicast address is generated by taking the last 24 bits of a unicast or anycast address and appending the address to the FF02::0:0:0:1:FF00::/104 prefix. In the case where the address 2000::1 is used, the address solicited node multicast address FF02::1:FF00:1 would be generated.
- IPv6 DAD is similar to the gratuitous ARP protocol used in IPv4 to detect duplicated IPv4 host addresses upon address assignment or connection of the host to the network. A node sends a neighbor solicitation (NS) message that requests the tentative address be used as the destination address to the Solicited-node multicast group.
- If the node receives a neighbor advertisement (NA) reply message, the tentative address is confirmed as being used by another node and the node will not use this tentative address for communication, following which manual assignment of an address by an administrator would be necessary.



## Summary

- What is the smallest condensed IPv6 value possible for the address  
2001:0DB8:0000:0000:0000:0000:032A:2D70
- How is it possible for an end station to independently generate an IPv6 address?

- The address 2001:0DB8:0000:0000:0000:0000:032A:2D70 can be condensed to 2001:DB8::32A:2D70. The double colon can be used to condense strings of zeroes that are often likely to exist due to the unfathomable size of the IPv6 address space, but may only be used once. Leading zero values can be condensed, however trailing zero values in address strings must be maintained.
- The 64-bit Extended Unique Identifier (EUI-64) format can be used to generate a unique IPv6 address. This is achieved by inserting two 'FF' and 'FE' octet values into the existing 48-bit MAC address of the interface, to create a unique 64-bit value that is adopted as the IPv6 address of the end station interface. Validation of the address as a unique value is confirmed through the Duplicate Address Detection (DAD) protocol.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# IPv6 Routing Technologies

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The changes to the address architecture have introduced the need for routing protocols that are capable of supporting IPv6. IPv6 Routing protocols include OSPFv3.
- The characteristics and operation of each of OSPFv3 generally reflects those used in OSPFv2, however contain some distinct differences that are required to be understood to support the implementation of IPv6 based routing protocols within an IPv6 founded network.

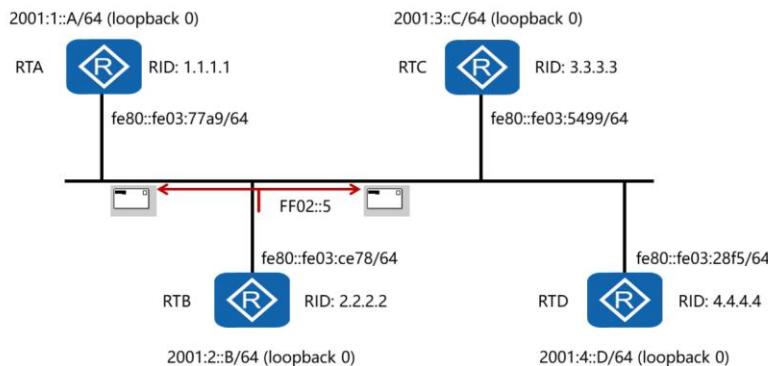


## Objectives

- Upon completion of this section, you will be able to:
  - Describe the characteristics and operation of OSPFv3.
  - Configure OSPFv3 routing protocols for IPv6.



## OSPFv3

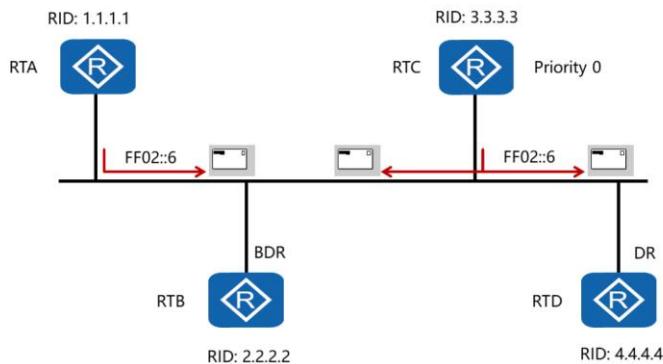


- OSPFv3 updates are sent to the All-SPF-Routers address ,that is FF02::5.
- Link-local addressing used by default to define the next-hop.

- OSPFv3 is the version of OSPF that is used over IPv6 networks, and in terms of communication, assumes that each router has been assigned link-local unicast addresses on each of the router's attached physical links. OSPF packets are sent using the associated link-local unicast address of a given physical interface as the source address. A router learns the link-local addresses of all other routers attached to its links and uses these addresses as next-hop information during packet forwarding. This operation is true for all physical links with the exception of virtual links which are outside the scope of this material.
- A reserved “AllSPFRouters” multicast address has been assigned the value FF02::5, reflecting the multicast address 224.0.0.5 used in OSPFv2, for which it should be noted that the two versions are not compatible. All routers running OSPFv3 should be prepared to receive packets sent to this address. Hello packets are always sent to this destination, as well as certain OSPF protocol packets during the flooding procedure.



## OSPFv3 Router ID

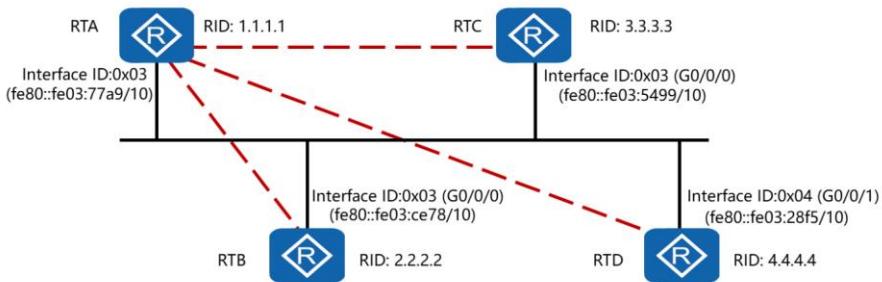


- Not based on any IP address, must be manually defined.
- Router ID continues to be used to support DR and BDR election.

- The router ID plays a prominent role in OSPFv3 and is now always used to identify neighboring routers. Previously, they had been identified by an IPv4 address on broadcast, NBMA (Non-Broadcast Multi-Access), and point-to-multipoint links. Each router ID (as well as the area ID) is maintained as a 32 bit dotted decimal value and cannot be configured using an IPv6 address.
- The router ID also continues to actively act as a tie breaker in the event that the priority associated with each OSPFv3 enabled router is equal. In such cases the Designated Router (DR) is identified as the router possessing the highest router ID. Hello messages sent will contain a router ID set to 0.0.0.0 if there is no Designated Router. The same principle applies for the Backup Designated Router, identified by the next highest router ID. A priority of 0 set on a router interface participating in OSPFv3 will deem the router as ineligible to participate in DR/BDR elections. Router Priority is only configured for interfaces associated with broadcast and NBMA networks.
- The multicast address “AllDRouters” has been assigned the value FF02::6, the equivalent of the 224.0.0.6 multicast address used in OSPFv2 for IPv4. The Designated Router and Backup Designated Router are both required to be prepared to receive packets destined to this address.



## OSPFv3 Per Link Behavior

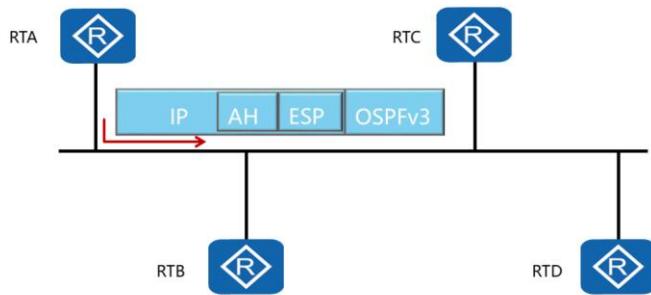


- OSPFv3 operates on the principle of per-link as opposed to the per-network or per-subnet concept used in IPv4.

- IPv6 refers to the concept of links to mean a medium between nodes over which communication at the link layer is facilitated. Interfaces therefore connect to links and multiple IPv6 subnets can be associated with a single link, to allow two nodes to communicate directly over a single link, even when they do not share a common IPv6 subnet (IPv6 prefix). OSPFv3 as a result operates on a per-link basis instead of per-IP-subnet as is found within IPv4. The term link therefore is used to replace the terms network and subnet which are understood within OSPFv2. OSPF interfaces are now understood to connect to a link instead of an IP subnet. This change affects the receiving of OSPF protocol packets, the contents of Hello packets, and the contents of network-LSAs.
- The impact of links can be understood from OSPFv3 capability to now support multiple OSPF protocol instances on a single link. Where separate OSPF routing domains that wish to remain separate but operate over one or more physical network segments (links) that are common to the different domains. IPv4 required isolation of the domains through authentication which did not provide a practical solution to this design requirement.
- The per-link operation also means the Hello packet no longer contains any address information, and instead it now includes an Interface ID that the originating router assigns to uniquely identify (among its own interfaces) its interface to the link.



## OSPFv3 Authentication



- OSPF authentication removed in OSPFv3, instead relying on the AH & ESP extension headers of IP for security.

- Authentication in OSPFv3 is no longer supported, and as such the authentication type and authentication (value) fields from the OSPF packet header no longer exist. As of IPv6, changes to the IP header have allowed for the utilization of security protocols that were only present in IPv4 as part of the IPsec suite, since initial TCP/IP development in the late 1960's was never designed with security in mind, since security of TCP/IP at the time was not foreseen as a future requirement.
- With the security improvements made to IPv6 and the utilization of IP extension headers in the protocol suite, OSPFv3 is capable of relying on the IP Authentication Header and the IP Encapsulating Security Payload to ensure integrity as well as authentication & confidentiality, during the exchange of OSPFv3 routing information.



## Enabling OSPFv3



```
[RTA] ipv6
[RTA] ospfv3
[RTA-ospfv3-1]router-id 1.1.1.1
[RTA-GigabitEthernet0/0/0]ipv6 enable
[RTA-GigabitEthernet0/0/0]ipv6 address fe80::1 link-local
[RTA-GigabitEthernet0/0/0]ospfv3 1 area 0.0.0.0
[RTA-LoopBack0]ipv6 enable
[RTA-LoopBack0]ipv6 address 2001:1::1/64
[RTA-LoopBack0]ospfv3 1 area 0.0.0.0
```

- Implementing OSPFv3 between peers requires, as with RIPng, that the router firstly be capable of supporting IPv6. Additionally the router must also enable the OSPFv3 protocol globally in the system view. Each interface should be assigned an IPv6 address. During the configuration of RIPng it was demonstrated how the interface can be automatically configured to assign a link local address. In this example an alternative and recommended method of link local address assignment is demonstrated. The link local IPv6 address is manually assigned and designated as a link local address using the *ipv6 <link local address> link-local* command. If the address associated with the physical interface is a global IPv6 unicast address, the interface will also automatically generate a link local address.
- In order to generate an OSPFv3 route, the example demonstrates assigning global unicast addresses to the logical loopback interface of each router. The physical and logical interfaces both should be associated with the OSPFv3 process and be assigned a process ID (typically the same ID unless they are to operate as separate instances of OSPFv3) and also be assigned to an area, which in this case is confined to area 0.
- As a reminder, OSPFv3 nodes rely on identification by neighboring routers through the use of the router ID, and therefore a unique router ID should be assigned for each router under the OSPFv3 protocol view following the configuration of the *ospfv3* command.



## Configuration Validation

```
[RTA]display ospfv3
Routing Process "OSPFv3 (1)" with ID 1.1.1.1
Route Tag: 0
Multi-VPN-Instance is not enabled
SPF Intelligent Timer[millisecs] Max: 10000, Start: 500, Hold: 2000
LSA Intelligent Timer[millisecs] Max: 5000, Start: 500, Hold: 1000
LSA Arrival interval 1000 millisecs
Default ASE parameters: Metric: 1 Tag: 1 Type: 2
Number of AS-External LSA 0. AS-External LSA's Checksum Sum 0x0000
Number of AS-Spaced Unknown LSA 0. AS-Spaced Unknown LSA's Checksum
Sum 0x0000
Number of FULL neighbors 1
Number of Exchange and Loading neighbors 0
.....
```

- Following the configuration of neighboring routers participating in OSPFv3, the display ospfv3 command is used to verify the operation and parameters associated with OSPFv3. Each router will show a running OSPFv3 process and a unique router ID. If the adjacency between neighbors is established, the number of FULL (state) neighbors will show a value greater than zero.



## Summary

- What is used to uniquely identify each neighboring node running OSPFv3?

- A 32-bit router ID is used to uniquely distinguish each node running OSPFv3.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## IPv6 Application Services DHCPv6

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The IPv6 architecture has led to the redesign of many aspects of network operation. One such design change involves Neighbor Discovery, which in itself now defines a means for Stateless Address Auto-configuration (SLAAC). DHCP for IPv6 (DHCPv6) includes a number of design changes that includes support for both SLAAC and stateful IPv6 addressing. DHCPv6 remains a client/server based application layer protocol, however includes a significant number of changes to align with the design aspects of IPv6. As such, DHCPv6 stateful and stateless implementations and characteristics are explained.

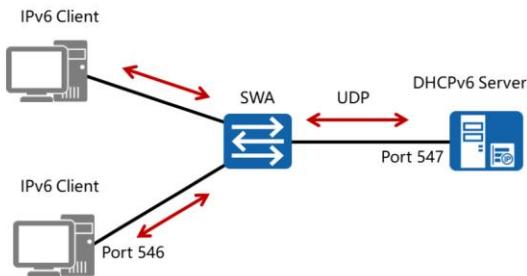


## Objectives

- Upon completion of this section, you will be able to:
  - Describe the features of DHCPv6.
  - Explain the stateful and stateless behavior of DHCPv6.
  - Successfully configure DHCPv6 services.



## DHCPv6

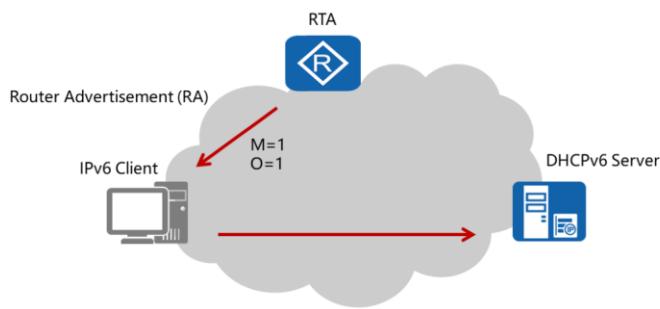


- Represents a stateful address auto-configuration protocol.
- UDP based communication between client and server.

- The Dynamic Host Configuration Protocol for IPv6 (DHCPv6) is a technology that dynamically manages and configures IPv6 addresses in a centralized manner. It is designed to assign IPv6 addresses and other network configuration parameters to hosts. DHCPv6 uses the client/server model. A client requests configurations such as the IPv6 address and DNS server address from the server, the server replies with requested configurations based on policies.
- In stateless address auto-configuration (SLAAC), routers do not record the IPv6 addresses of the hosts, therefore stateless address auto-configuration has poor manageability. In addition, hosts configured with stateless address auto-configuration cannot obtain other configuration parameters such as the DNS server address. ISPs do not provide instructions for automatic allocation of IPv6 prefixes for routers. Users therefore need to manually configure IPv6 addresses for devices during IPv6 network deployment.
- As a stateful protocol for configuring IPv6 addresses automatically, DHCPv6 solves this problem. During stateful address configuration, the DHCPv6 server assigns a complete IPv6 address to a host and provides other configuration. Parameters such as the DNS server address and domain name. A relay agent may be used to forward DHCPv6 packets, however lies outside of the scope of this material. The DHCPv6 server binds the IPv6 address to a client, improving overall network manageability.
- Clients and servers exchange DHCP messages using UDP. The client uses a link-local address, determined through other mechanisms for transmitting and receiving DHCP messages. Clients listen for DHCP messages on UDP port 546, whilst servers (and relay agents) listen for DHCP messages on UDP port 547.



## Stateful Addressing



- RA contains managed (M) and other (O) configuration flags.
- Stateful addressing (DHCPv6) used where flags are set to '1' .

- Prior to the allocation of addresses, it should be clearly understood that an IPv6 node (client) is required to generate a link-local address and be successfully evaluated by the Duplicate Address Detection (DAD) process. Following this, a link router discovery process is involved, for which the IPv6 client node broadcasts a Router Solicitation (RS) message, and the link router responds with a Router Advertisement (RA) message after receiving the RS message.
- Contained within the RA message are numerous fields containing configuration parameters for the local network. One field in particular referred to as the Autoconfig Flags field, is an 8 bit field that contains two specific bit values to determine the auto-configuration process for the local network. A "managed address configuration flag" (M) is a 1 bit value that is for defining whether stateful address configuration should be used, commonly applied in the presence of a DHCPv6 server on the local network. Where the value is set to 1, stateful addressing should be implemented, meaning the IPv6 client should obtain IPv6 addressing through stateful DHCPv6.
- The other stateful configuration flag (O) represents the second flag bit value in the Autoconfig Flags field, and defines whether other network configuration parameters such as DNS and SNTP (for time management servers) should be determined through stateful DHCPv6. RFC2462 defines that where the M bit is true (a value of 1), the O bit must also be implicitly true, however in practice the M bit and the O bit may be defined interchangeably to support stateless addressing services in DHCPv6, in which an IPv6 address is not assigned but configuration parameters are.
- It should also be noted that the managed address flag and other configuration flag is managed through VRP on the router, and is not set in the RA message by default. In order to set these flags, the commands `ipv6 nd autoconfig managed-address-flag` and `ipv6 nd autoconfig other-flag` should be configured on the gateway responsible for generating RA messages.



## Enabling DHCPv6 Communication

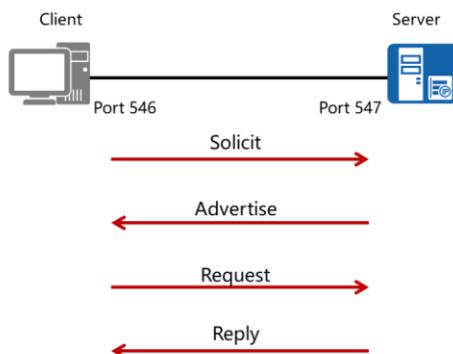


- Link-local addresses are used as source address by clients, and DHCP servers reached via the multicast address FF02::1:2.

- Client nodes initiated on a network supporting stateful addressing may be serviced by one or more DHCPv6 servers. The IPv6 client uses a link-local address assigned to the interface for which it is requesting configuration information as the source address in the header of the IPv6 datagram.
- The multicast address FF02::1:2 is a reserved multicast address that represents "All\_DHCP\_Relay\_Agents\_and\_Servers" , and is used by a client to communicate with neighboring servers. All servers (and relay agents) are members of this multicast group. For any client sending a DHCP message to the All\_DHCP\_Relay\_Agents\_and\_Servers address, it is expected that the client send the message through the interface for which configuration information is being requested, however exceptions may occur to this rule where two interfaces on the client are associated with the same link, for which it is possible for the alternative interface to be used. In either case the link local address of the forwarding interface must be used as the source address.



## Assigning IPv6 Addressing

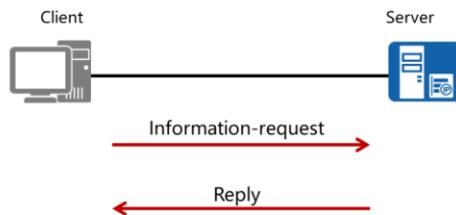


- Discovery of servers and assignment of IPv6 addresses & configuration parameter relies on a set of four messages.

- Obtaining stateful addressing and other parameters from a DHCPv6 server requires a series of messages be sent. A client initially sends a solicit message to locate servers, from which addressing and configuration parameters can be received.
- Following the solicit message, a DHCPv6 server supporting the link will generate an advertise message in response to the solicit message, that indicates to the client, the IPv6 address of the server, providing the required DHCPv6 service. The client is then capable of using this IPv6 address to reference the DHCPv6 server and generate a request message. Where multiple servers respond to the solicit message, the client will need to decide which DHCPv6 server should be used, typically defined by a server preference value defined by the DHCPv6 administrator on the server, and carried in the advertise message. Additionally the server may carry options including a server unicast option which enables the client to use the IPv6 address of the DHCPv6 server to transmit further correspondence with this server as unicast messages.
- The request message is transmitted to the selected DHCP server to request configuration parameters and one or multiple IPv6 addresses to be assigned. Finally the DHCPv6 server responds with a Reply message that contains the confirmed addresses and network configuration parameters.



## Stateless Configuration Information

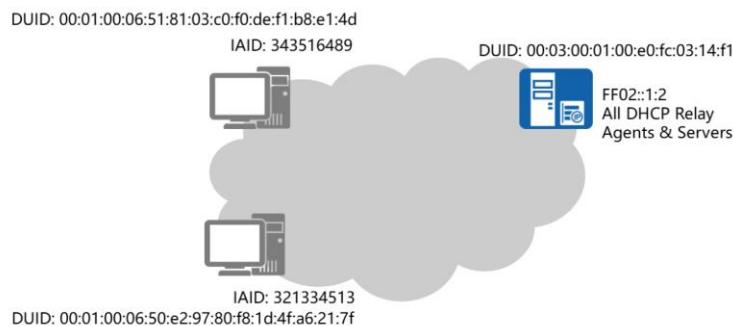


- Information-request used when IPv6 addressing not required.
- Reply used to deliver configuration parameters.

- DHCP may also be employed to support stateless configuration in the event where a host is capable of retrieving IPv6 addressing information through stateless configuration means, and requires only specific configuration parameters from the DHCPv6 server. In such events Information-request messages are generated, and sent by clients to a server to request configuration parameters. The client is able to obtain configuration information such as server addresses and domain information, as a list of available configuration parameters, using only a single message and reply that is exchanged with a DHCP server.
- The Information-Request message is sent to the “All\_DHCP\_Relay\_Agents\_and\_Servers” multicast address following which servers respond with a Reply message containing the configuration information for the client. Since no dynamic state is being maintained (i.e. in the form of IPv6 address assignment) the allocation of configuration information is understood to be stateless.



## DHCP Unique Identifier (DUID)

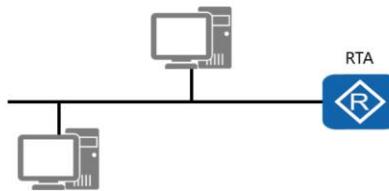


- Unique identifier of clients & servers in the DHCP community.
- Parameters bound to each DUID using Identity Associations (IA).

- A DHCP Unique Identifier (DUID) is a value that is used to distinguish between each client and DHCP server, for which only one DUID is present in each case. Clients may have one or multiple interfaces for which each will be assigned an IPv6 address along with other configuration parameters and is referenced using an Identity Association Identifier (IAID). These are used together with DUID to allow DHCPv6 servers to reference a client and the interface address/configuration parameters that should be assigned.
- In the case of each client, the DUID will be used to identify a specific DHCP server with which a client wishes to communicate. The length of the DUID value can vary from anywhere in the range of 96bits (12 bytes) to 160 bits (20 bytes), depending on the format that is used. Three such formats exist, using either the link-layer (DUID-LL) address, a combination of the link-layer address and enterprise number (DUID-EN), a vendor assigned value at the point of device manufacture, or a combination of the link-layer address and a timestamp value (DUID-LLT) generated at the point of DUID creation in seconds from midnight Jan 1st 2000 (GMT), modulo 232.
- The initial 16 bit values (00:01) represent the format used, where "00:01" denotes the DUID-LLT format, "00:02" the DUID-EN format and "00:03" the DUID-LL format. In the case of the DUID-LL and DUID-LLT formats, the 16 bits immediately after represent the hardware address based on IANA hardware type parameter assignments, with 00:01 representing Ethernet (10Mb) and 00:06 defining IEEE 802 network standards. A time stamp follows in the DUID-LLT format and finally the link layer address value. For DUID-LL only the link layer address follows.



## Setting the DHCP DUID



```
[RTA] dhcpv6 duid 11
```

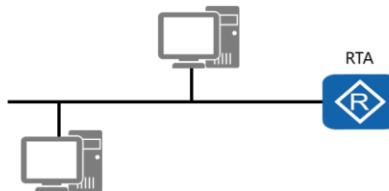
Warning: The DHCP unique identifier should be globally-unique and stable. Are you sure to change it? [Y/N]y

- Enables assignment of either the DUID-LL or DUID-LLT format.
- The DUID-LL format is assigned by default.

- The DUID format can be assigned through the `dhcpv6 duid` command, for which either the DUID-LL or DUID-LLT format can be applied. The DUID-LL format is applied by default. For the DUID-LLT, the timestamp value will reference the time from the point at which the `dhcpv6 duid llt` command is applied. The `display dhcpv6 duid` command can be used to verify the current format based primarily on the length of the DUID value, as well as the DUID value itself.



## IPv6 Address Pool



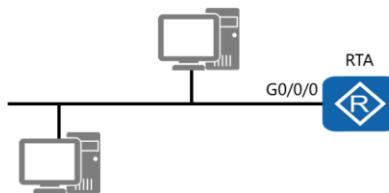
```
[RTA] dhcipv6 pool pool1
[RTA-dhcipv6-pool-pool1] address prefix 3000::/64
[RTA-dhcipv6-pool-pool1] excluded-address 3000::1
[RTA-dhcipv6-pool-pool1] dns-server 3001::1
[RTA-dhcipv6-pool-pool1] dns-domain-name huawei.com
```

- DHCPv6 parameters are assigned for each address pool.

- The implementation of stateful addressing requires that an address pool be defined with a typical address prefix defined for the given range, as well as pool specific parameters. The example demonstrates how a pool is created with the defining of a pool and associated pool name, as well as the address prefix from which the range of host addresses will be allocated.
- Excluded addresses refer to addresses that comprise of the pool range however should not be allocated through DHCP since they are commonly used for other applications such as the interface address of the DHCP server. Additional configuration parameters will also be specified for a given pool with examples such as server addresses and domain names being specified for parameter allocation to DHCPv6 clients.



## Enable DHCPv6 Server



```
[RTA] ipv6
[RTA] dhcp enable
[RTA] interface GigabitEthernet 0/0/0
[RTA-GigabitEthernet0/0/0] ipv6 enable
[RTA-GigabitEthernet0/0/0] ipv6 address 3000::1/64
[RTA-GigabitEthernet0/0/0] dhcpv6 server pool1
```

- Address pool is associated with the DHCPv6 server interface.

- A created DHCPv6 pool is required to be associated with an interface through which it will service DHCP clients. An IPv6 address is assigned to the interface of the DHCPv6 server and the interface then associated with the address pool. In this case the excluded address value has been used to represent the interface address in order to ensure no attempts are made by the DHCPv6 server to assign the interface address to a DHCP client.



## Displaying DHCPv6 Information

```
<RTA>display dhcpv6 pool
DHCPv6 pool: pool1
    Address prefix: 3000::/64
        Lifetime valid 172800 seconds, preferred 86400 seconds
        2 in use, 0 conflicts
    Excluded-address 3000::1
    Information refresh time: 86400
    DNS server address: 3001::1
    Domain name: huawei.com
    Conflict-address expire-time: 172800
    Active normal clients: 2
```

- Configured pools, pool based parameters, and client activity are referenced under the display dhcp pool command.

- The resulting configuration of the DHCPv6 server can be clarified through the display dhcpv6 pool command, from which point the defined pool(s) can be identified and the address prefix associated with the pool determined. Additional information such as the lifetime can be viewed, for which the default lifetime of a leased address is 86400 seconds, or 1 day and can be reconfigured as necessary using the information-refresh command under the dhcpv6 pool <pool-name> view. Where active clients have leased addresses from the DHCP server, the related statistics can be found here.



## Summary

- Which DUID formats are currently supported within VRP?
- If the M and O bits of a Router Advertisement (RA) are set to 1, what action is taken by the client?

- The AR2200 series router is capable of supporting Link Layer (DUID-LL) and Link Layer with Time (DUID-LLT) formats.
- When a router advertisement is received containing M and O bit values set to 1, the client will perform discovery of a DHCPv6 server for stateful address configuration. The configuration will include an IPv6 address and also other configuration information such as prefixes and addressing of servers providing services such as DNS.

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## MPLS Basic Principle

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- The course introduces the MPLS architecture that contains two planes, MPLS encapsulation mode and label format. The process of MPLS data forwarding is the key point of the course.



## Objectives

- Upon completion of this course, you will be able to:
  - Describe the background and typical applications of MPLS.
  - Describe the basic concepts and working principles of MPLS.
  - Describe the process of MPLS data forwarding.



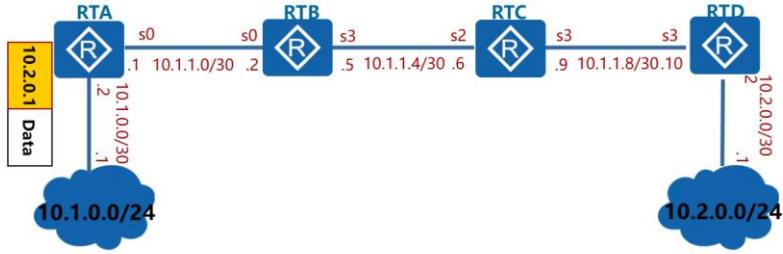
# Contents

1. MPLS Overview
2. MPLS Basic Principle



## Traditional IP forwarding

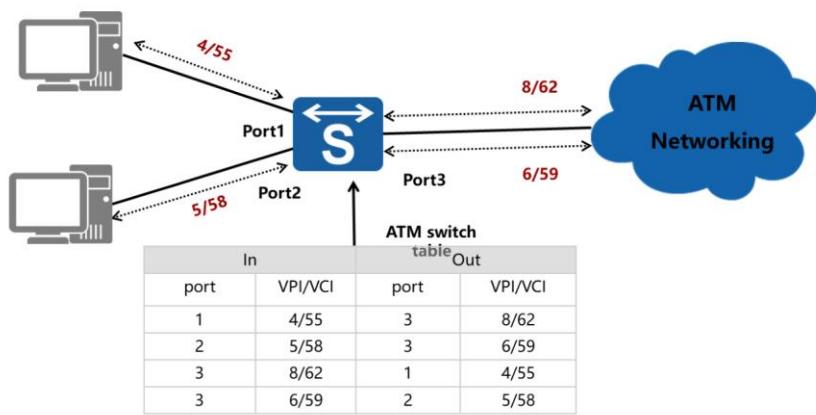
Network	Nexthop	Network	Nexthop	Network	Nexthop	Network	Nexthop
10.1.0.0/24	10.1.0.2	10.1.0.0/24	10.1.1.1	10.1.0.0/24	10.1.1.5	10.1.0.0/24	10.1.1.9
10.1.0.1/32	10.1.0.1	10.1.1.0/30	10.1.1.2	10.1.1.0/30	10.1.1.5	10.1.1.0/30	10.1.1.9
10.1.1.0/30	10.1.1.1	10.1.1.1/32	10.1.1.1	10.1.1.4/30	10.1.1.6	10.1.1.4/30	10.1.1.9
10.1.1.2/32	10.1.1.2	10.1.1.4/30	10.1.1.5	10.1.1.5/32	10.1.1.5	10.1.1.8/30	10.1.1.10
10.1.1.4/30	10.1.1.2	10.1.1.6/32	10.1.1.6	10.1.1.8/30	10.1.1.9	10.1.1.9/32	10.1.1.9
10.1.1.8/30	10.1.1.2	10.1.1.8/30	10.1.1.6	10.1.1.10/32	10.1.1.10	<b>10.2.0.0/24</b>	<b>10.2.0.2</b>
<b>10.2.0.0/24</b>	<b>10.1.1.2</b>	<b>10.2.0.0/24</b>	<b>10.1.1.6</b>	<b>10.2.0.0/24</b>	<b>10.1.1.10</b>	10.2.0.1/32	10.2.0.1



- In the traditional IP forwarding, physical layer receives a packet from a port on the router, then sends to data link layer.
- Data link layer removes link layer encapsulation, then sends to corresponding network layer according to protocol field of the packet.
- The network layer will check whether the packet is sent to this device, if it is, it will remove network layer encapsulation, and send to its up-level protocol. If it is not, it will lookup routing table according to packet's destination IP address, if the route is matched, the packet will be sent to data link layer of the corresponding port, after encapsulated by data link layer, it will be transmitted. If it is not matched, the packet will be discarded.
- The traditional IP forwarding adopts hop-by-hop forwarding, each router that packet passed through should implement the process (as the figure shows, RTA receives a data packet whose destination IP address is 10.2.0.1, RTA will lookup routing table and forward according to matched route item, RTB, RTC, RTD will also do like this), So the efficiency is low. And all the routers need to know all routes in the entire network or default route.
- Besides, the traditional IP forwarding is connectionless oriented, so it is hard to deploy Qos.

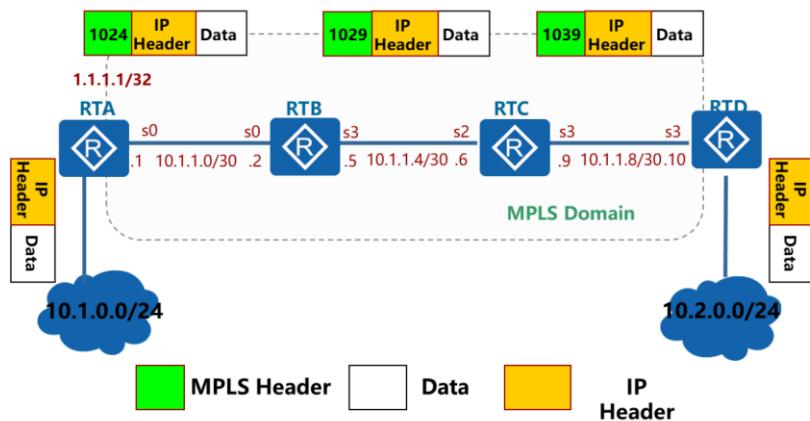


## ATM cell forwarding





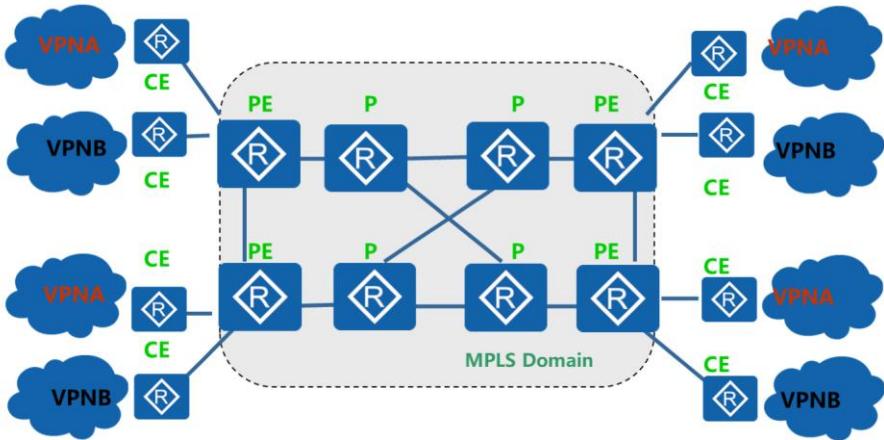
# MPLS label forwarding



- MPLS is a kind of label forwarding technology, it adopts connectionless control plane and connection oriented data plane, connectionless control plane implements routing transmission and label distribution, connection oriented data plane implements packet transmission along LSP (label switch path) established before.
- In MPLS network domain, router does not need to analyze every packet's destination IP address, just forwards by label that added before the IP header (as the figure shows that RTB receives labeled packet from RTA, then forwards by label, RTC is similar). Comparing to traditional IP forwarding, MPLS label forwarding greatly improves forwarding efficiency.



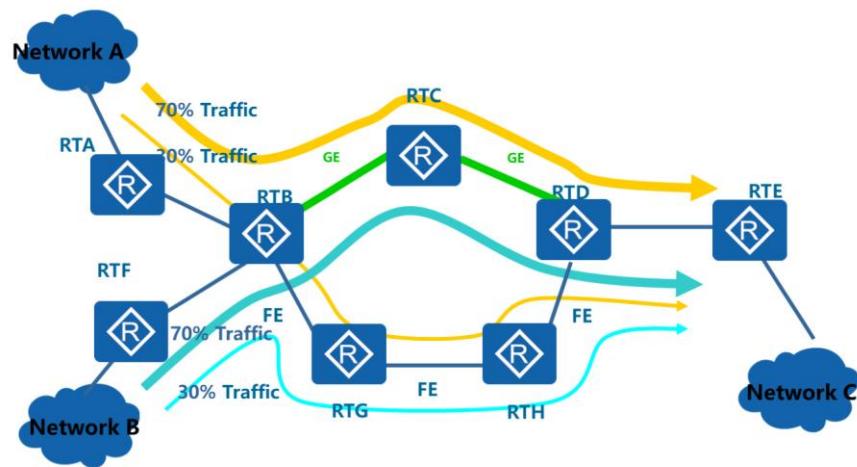
## MPLS VPN Application



- However, with the development of ASIC technology, routing lookup speed is not bottleneck of network development any more. Improving forwarding speed is no longer the obvious advantage of MPLS.
- MPLS integrates the advantage of the two forwarding technologies, powerful layer 3 routing function of IP network and high efficiency forwarding mechanism of traditional layer 2 network, its forwarding plane adopts connection oriented, it is very similar to layer 2 network forwarding method in existence.
- It makes MPLS easy to implement seamless combination of IP and ATM, frame relay and other layer 2 network, and provide better solution for TE (Traffic Engineering), VPN (Virtual Private Network), QoS (Quality of Service) and other applications.
- VPN based on MPLS can combine different embanchment of private network, form a uniform network, VPN based on MPLS also supports communication control between different VPN. As the figure shows, CE is user edge device; PE is service provider edge router, which is located in backbone network. P is backbone router in the service provider network, it does not directly connect with CE. VPN data is transmitted along LSP (label switch path) encapsulated with MPLS label.



## MPLS TE Application



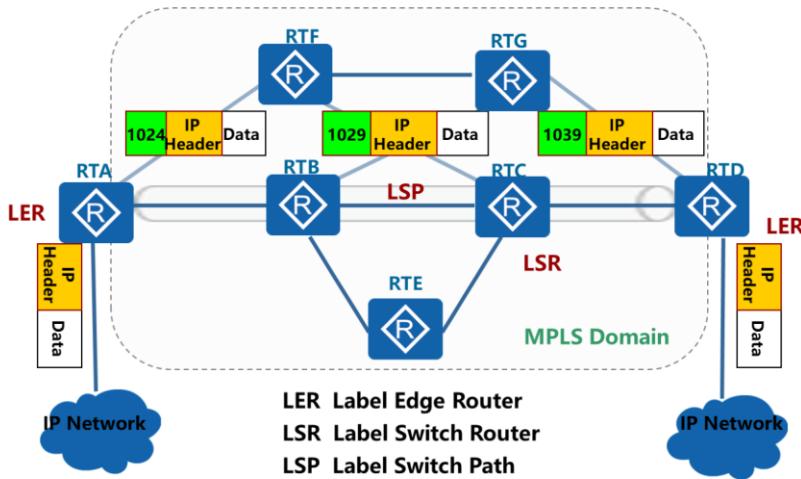


# Contents

1. MPLS Overview
2. MPLS Basic Principle
  - 2.1 MPLS control plane and forwarding plane
  - 2.2 MPLS Label Format
  - 2.3 MPLS Forwarding Process



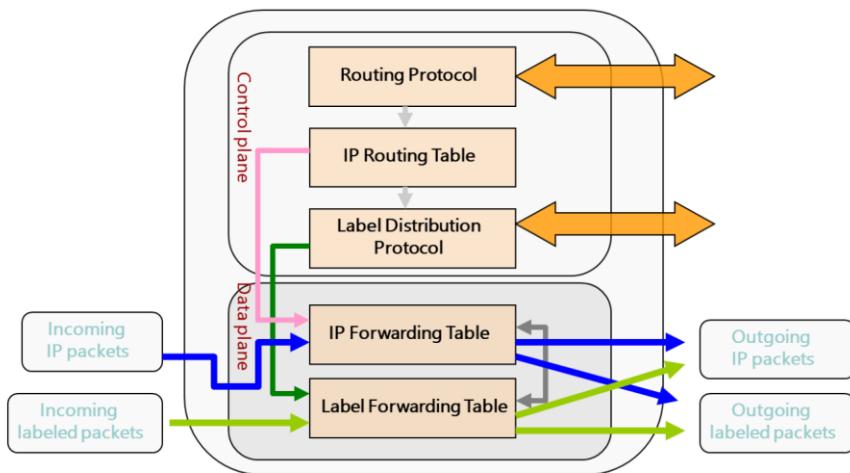
## MPLS network model



- The typical structure of MPLS network is shown in this slide: the router and ATM switch located inside of MPLS domain are called LSR, router and ATM switch located at the edge of MPLS domain that used to connect IP network or other kinds of network are called LER.
- In IP network, it implements traditional IP forwarding; in MPLS domain, it implements label forwarding.
- Both of LER and LSR have the ability of label forwarding, but they are located in different position, the packet processing is different. LER's charge is to receive IP packet from IP network and insert label into the packet, then transmit it to LSR, whereas, its charge is also to receive labeled packet from LSR and remove label, transmit it to IP network; LSR's charge is to forward according to the label.
- The path that packet passes through in MPLS domain is called Label Switch Path (LSP), this path is already confirmed and established by kinds of protocols before packet forwarding, packet will be transmitted along the specified LSP.



# MPLS control plane and forwarding plane



- MPLS network forwards packet according to the label. But how does label be generated? What mechanism does MPLS adopt to implement data forwarding?
- MPLS includes two plane: control plane and data plane.
  - Control plane' s charge is to generate and maintain routing information and label information. Data plane' s charge is conventional IP packet forwarding and labeled packet forwarding. In control plane, routing protocol module is used to transmit routing information, generate routing table; label distribution protocol is used to complete exchange of label and establish label switch path.
  - Data plane includes IP forwarding table and label forwarding table, when receiving conventional IP packets, if it is conventional IP forwarding, it should lookup routing table and forward, if it is label forwarding, it should forward by the label forwarding table; when receiving labeled packets, if it needs to forward by label, it should forward by label forwarding table, if it needs to transmit to IP network, it should remove label and forward by IP routing table.



# Contents

1. MPLS Overview
2. MPLS Basic Principle
  - 2.1 MPLS control plane and forwarding plane
  - 2.2 MPLS Label Format
  - 2.3 MPLS Forwarding Process



## Frame mode MPLS

- MPLS has two encapsulation modes: Frame mode and cell mode (ATM uses the MPLS cell encapsulation mode, which is not involved in this course). In frame encapsulation mode, an MPLS label header is added between the Layer 2 header and Layer 3 header of a packet. Ethernet and PPP use this encapsulation mode.



Layer 2 frame format



MPLS frame mode encapsulation

- MPLS labels are used to transmit MPLS information. Routers exchange labels to transmit data on the established label forwarding paths.



## MPLS Header

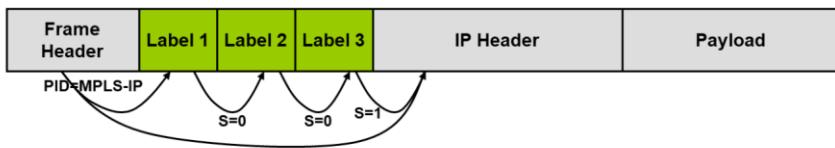


- The total length of MPLS header is 4bytes (32bits)
- The length of Label field is 20bits
- The length of EXP (Experimental Use) field is 3bits
- The length of S (Bottom of Stack) field is 1bit
- The length of TTL field is 8bits

- The length of MPLS header is 32 bits, it includes 20-bit label field, this field is used for data forwarding; 3-bit EXP is used to carry precedence of IP packet; 1-bit S is at the bottom of stack and is used to indicate whether it is the last label (MPLS label can be multiple nesting); 8-bit TTL, its function is similar to TTL of IP header, it is used to prevent data from looping.



## MPLS Label Nesting



- PID indicates the types of packet follows Frame Header
  - Ethernet: 0x8100=IPv4, 0x8847=Unicast MPLS packet, 0x8848=Multicast MPLS packet
  - PPP: 0x8021=IPv4, 0x8281=Unicast MPLS packet, 0x8283=Multicast MPLS packet
- S indicates whether it is the last label
- Applications of label nesting
  - MPLS VPN
  - MPLS TE

- The protocol filed PID in layer 2 header specifies that payload starts with packet with label encapsulated or IP header. For example, in Ethernet protocol, PID=0x8847 identifies that the frame payload is a multicast MPLS packet. PID=0x8848 identifies that the frame payload is a unicast MPLS packet. PID=0x8000 identifies that the frame payload is a unicast IP packet. In PPP protocol, PID=0x8281 identifies that the frame payload is a unicast MPLS packet. PID=0x8283 identifies that the frame payload is a multicast MPLS packet.
- S bit in MPLS header indicates whether the next header is another label or a layer 3 IP header.
- Usually MPLS only allocates one label for a packet. But some advanced applications of MPLS use multiple labels. For example, MPLS VPN will use 2 layers of labels (in complex situation, it even uses 3 layers of labels), out-label is used for public network forwarding, in-label is used to indicate that which VPN the packet belongs to; MPLS TE also uses two or more labels, the outmost label is used to indicate TE tunneling, in-label indicates the destination of packet.
- Note: The Label1, Label2, Label3 all mean 4 Bytes MPLS header in last slide, it includes 20-bit label information.



# Contents

1. MPLS Overview
2. MPLS Basic Principle
  - 2.1 MPLS control plane and forwarding plane
  - 2.2 MPLS Label Format
  - 2.3 MPLS Forwarding Process



## FEC and NHLFE

- FEC (Forwarding Equivalence Class) is a set of data flows with certain common characteristics. These data flows are processed by the LSR in the same way during the forwarding process.
- FEC can be classified based on the address, service type, and QoS. For example, in the traditional IP forwarding that uses the longest matching algorithm, all packets to the same route are an FEC.
- NHLFE (Next Hop Label Forwarding Entry) : The NHLFE is used for label forwarding. It contains the following basic information:
  - Next hop of the packet
  - How to perform a tag operation (including pushing a new tag, popping a tag, and swapping the original tag with a new tag).
  - The NHLFE may also contain other information, such as the link layer encapsulation used for sending packets.



## Example of FEC and NHLFE

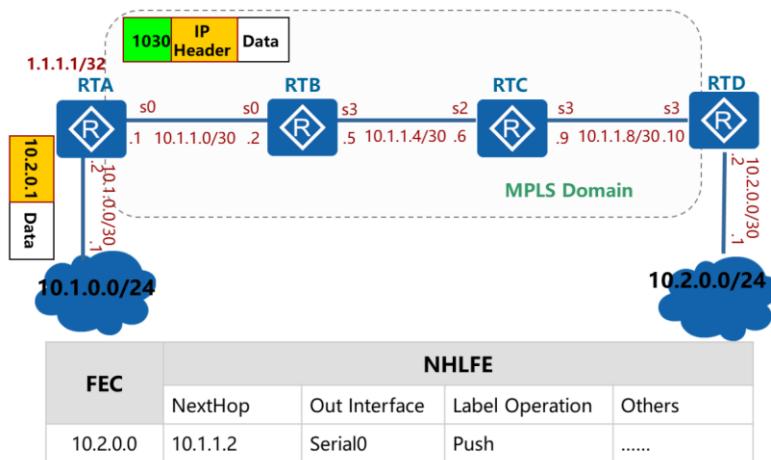
- **FEC: Forwarding Equivalence Classes**
- **NHLFE: Next Hop Label Forwarding Entry**

```
<RTA>display mpls lsp include 10.2.0.0 24 verbose
-----
LSP Information: LDP LSP
-----
No : 1
VrfIndex :
Fec : 10.2.0.0/24
Nexthop : 10.1.1.2
In-Label : NULL
Out-Label : 1030
In-Interface :
Out-Interface : Serial0
LspIndex : 10249
Token : 0x22005
LsrType : Ingress
Outgoing token : 0x0
Label Operation : PUSH
Mpls-Mtu : 1500
TimeStamp : 822sec
```

- FEC (Forwarding equivalence class) means a group of IP packets which are forwarded in the equipollence method, for example, a group of IP packets with same destination IP prefix will be allocated a unique label. In this case, the packet that destination IP prefix is 10.2.0.0/24 belongs to a FEC, the label allocated for this FEC is 1030.
- NHLFE is used when forwarding a labeled packet, It contains the following information:
  - 1. the packet's next hop;
  - 2.the operation to perform on the packet's label stack (it contains pushing new label, popping label, replacing the original label with new label). It may also contain other information, such as the data link encapsulation to use when transmitting the packet. In this case, next hop is 10.1.1.2, label operation is “push” .



## MPLS Forwarding - Ingress LER (RTA)

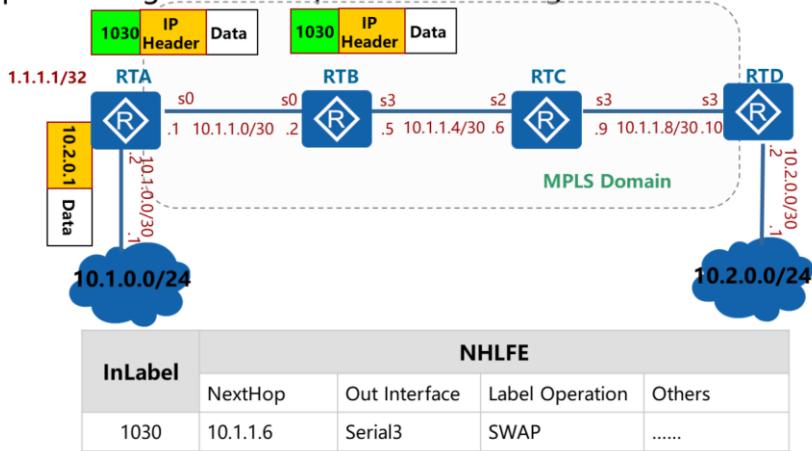


- FEC represents the same kind of packets, NHLFE contains next hop, label operation and other information. Only associating FEC with NHLFE, it can implement particular label forwarding for same kind of packets, FTN can implement this function, FTN (FEC-to-NHLFE) indicates the mapping for an FEC to NHLFE, if there are multiple cost-equal paths, one FEC maybe map to multiple NHLFE.
- When a IP packet enters MPLS domain, ingress LER (RTA) will analyze packet, determine which label to encapsulate packet according to packet characteristic (generally by prefix analysis of destination IP address), and determine to transmit to which next hop from which interface.
- On the ingress, the NHLFE table is queried to guide packet forwarding.



## MPLS Forwarding - LSR (RTB)

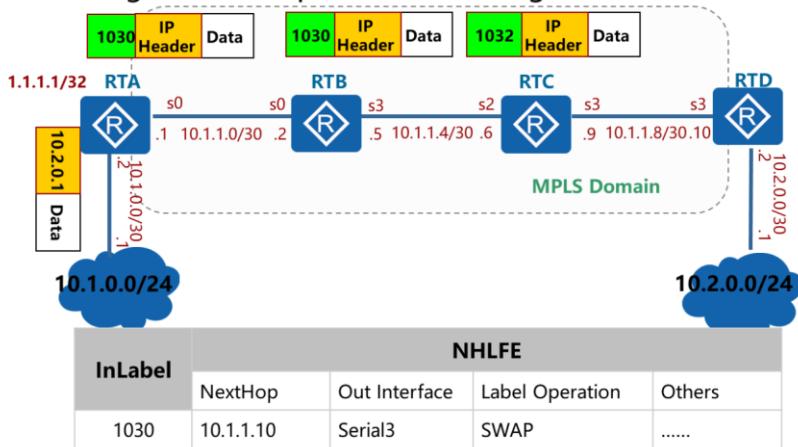
- On the transit node, the ILM(Incoming Label Map) table and NHLFE table are queried to guide MPLS packet forwarding.





## MPLS Forwarding - LSR (RTC)

- On the transit node, the ILM(Incoming Label Map) table and NHLFE table are queried to guide MPLS packet forwarding.

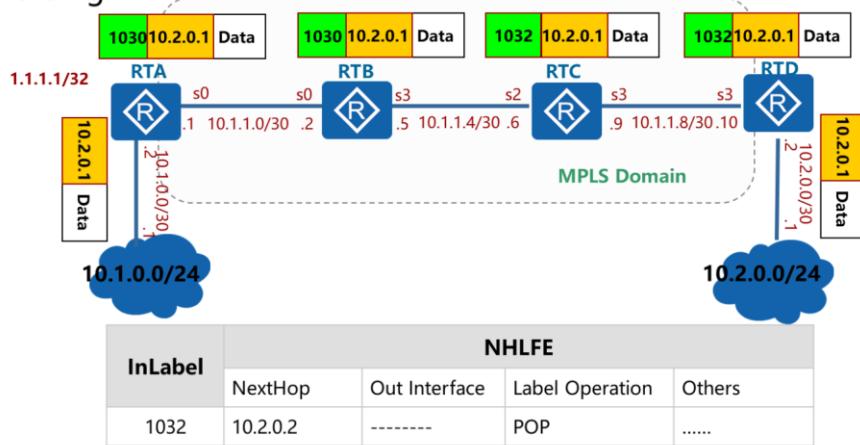


- Similar to RTB, when RTC receives message with label 1030, it forwards packet by label, and uses new outgoing label to swap original label.
- In this case, RTC uses outgoing label 1032 to swap incoming label, then transmits packet from outgoing interface Serial3, the next hop is 10.1.1.10.



## MPLS Forwarding - Egress LER (RTD)

- On the egress node, the ILM table is queried to guide MPLS packet forwarding.





## Quiz

1. Which field in the MPLS header is used to identify the stack bottom label ? ( )
  - A. Label
  - B. EXP
  - C. S
  - D. TTL
2. Which of the following actions can be performed when packets are forwarded based on MPLS labels? ( )
  - A. Push
  - B. Pop
  - C. Swap
  - D. Switch



## Summary

- Background of MPLS
- Basic Concepts of MPLS
- MPLS packet format
- MPLS Forwarding Principle

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



## SR Basic Principle

Copyright © 2019 Huawei Technologies Co., Ltd. All rights reserved.



## Foreword

- As a key technology of SDN, SR (Segment Routing) is also a key technology for simplified protocols advocated by IP networks. It is widely used in IP networks. The course will analyze the SR technology.



## Objectives

- Upon completion of this course, you will be able to:
  - Describe the advantages of SR.
  - Describe the basic concepts and working principles of SR.

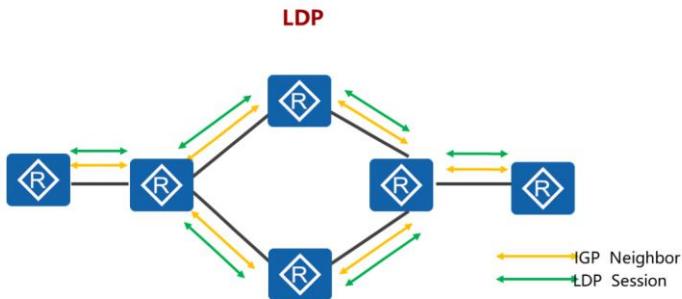


# Contents

1. SR Technical Background
2. SR Principle Analysis



## SR Background - Problems Faced by LDP

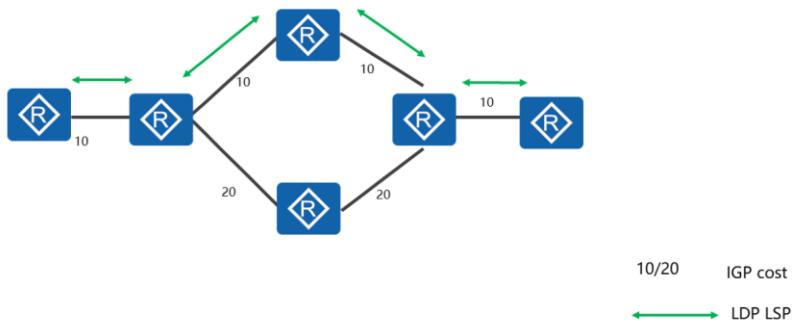


- LDP depends on IGP and uses local label switching to support ECMP.
- Disadvantages: LDP has 11 types of protocol packets, which greatly increase link bandwidth consumption and device CPU usage.

- LDP (Label Distribute Protocol): The label distribution protocol allocates labels based on destination addresses and uses label forwarding instead of IP forwarding. The advantage of LDP: It isolates public network routes, selects forwarding paths based on optimal routes, and supports ECMP. And the configuration is simple.
- Different types of services, such as voice, data, video, and VR, have different requirements on the network. The bandwidth requirements are increasing, and the network scale is increasing explosively. The traditional MPLS technology requires a dedicated label distribution protocol. Labels need to be allocated to each LSP, which occupies a large amount of resources. Status maintenance protocol packets occupy a large amount of bandwidth. It needs to be synchronized with the IGP protocol. The deployment and maintenance are complex and the scalability is poor. This network operation mode cannot meet the requirements of service providers to rapidly deploy network services on demand. In addition, the OPEX (Operating Expense) increases linearly with scale.



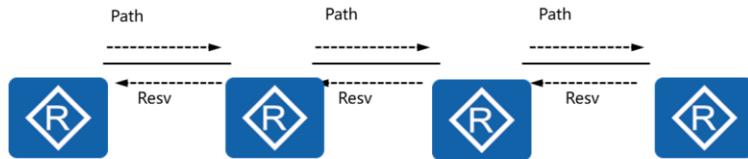
## SR Background - Problems Faced by LDP



- LDP supports only the shortest IGP path (minimum cost) for path calculation. TE(Traffic Engineering) is not supported.



## SR Background - Problems Faced by RSVP

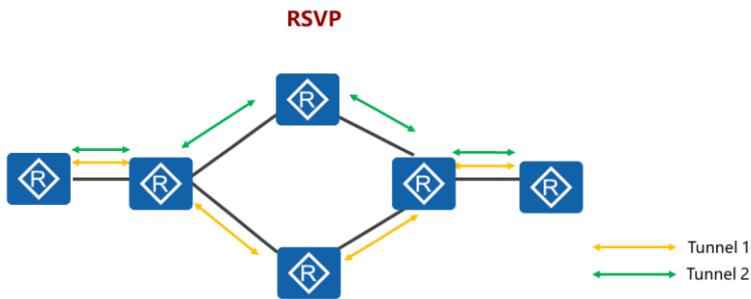


- RSVP creates an end-to-end LSP hop by hop based on the CSPF path calculation result. The label is a local label.
- Hop-by-hop, traversing nodes maintain the status of the tunnel. Even if SDN is used, the tunnel status needs to be maintained.
- RSVP TE configuration is complex.
- Complex ECMP implementation
- Eight types of RSVP protocol packets occupy network bandwidth and CPU processing.

- RSVP:Resource Reservation Protocol. RSVP-TE is introduced to solve the problem that the traditional IP network can forward only the optimal path and the path cannot be planned. RSVP-TE brings many benefits, such as explicit path planning, bandwidth resource reservation, and multiple protection schemes.
- RSVP relies on IGP to maintain its neighbor and link status, complicates the control plane, and complicates network maintenance and fault location.
- RSVP-TE needs to establish multiple tunnels to implement the ECMP function.
- In the LSP status maintenance, PATH and RESV need to be updated continuously. After services increase, performance problems occur at the intermediate point (services are overlapped).In addition, the configuration quantity of the RSVP-TE tunnel is always criticized by the operator. An average of eight tunnels need to be configured for each tunnel.



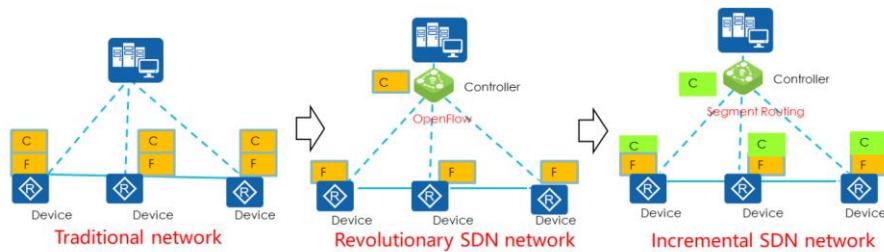
## SR Background - Problems Faced by RSVP



- RSVP is complex in load balancing and needs to be configured with multiple tunnels, which brings huge workload to configuration and maintenance.



# SR Generation Background -SR Technology



- The protocol evolution is slow. It takes several years from requirement submission to version release.
- The OPEX increases linearly with the scale.
- Lack of flexible deployment and use modes cannot meet service providers' requirements for fast network service deployment.

- Network faults can be rectified only when the controller is faulty. The reliability is poor.
- The number of flow tables on a large-scale network is large, and the flow table delivery rate has a performance bottleneck.
- It is a subversion of traditional networks and is not supported by equipment vendors and carriers.

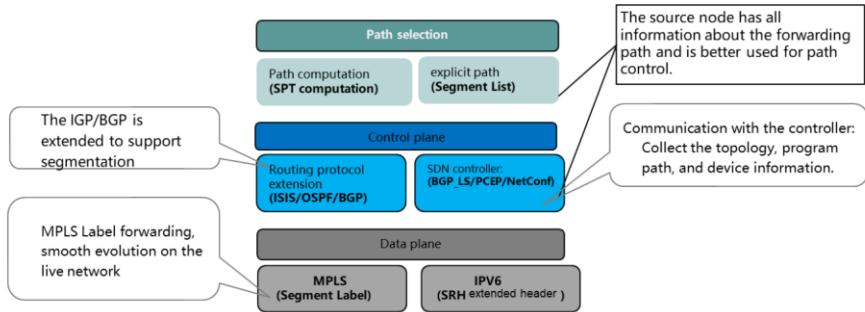
- The existing protocols can be expanded to achieve better smooth evolution.
- Provides the balance between centralized control and distributed deployment.
- Uses the source routing technology to provide fast interaction between the network and upper-layer applications.

- C:Control
- F:Forwarding
- SDN (Software Defined Network) is a new network innovation architecture proposed by the Stanford University clean slate Research Group. The core technology separates the control plane from the data plane to implement flexible control of network traffic and provide a good platform for the innovation of core networks and applications.
- The revolutionary SDN network uses a centralized control architecture. The control functions (such as route calculation) of network devices are centralized on one controller, and the forwarding table is generated and delivered by the controller to the device. Device functions are simplified and are only responsible for forwarding. OpenFlow is a control interface between the Controller and devices.
- The incremental SDN network expands the existing network. The existing devices evolve to the SDN. Some control functions are reserved on the device. Some functions that require centralized control are processed on the controller, which emphasizes the smooth evolution capability of the device.



## SR Background -SR Technical Framework

- SR is a technical solution of packet forwarding provided by adding a series of segment identifiers to a packet only on the source (an explicit path loading node) node





# Contents

1. SR Technical Background
2. SR Principle Analysis
  - 2.1 Basic Concepts of SR
  - 2.2 SR Working Principle



## Basic Concepts of SR

- **SR Domain:** Segment Routing Domain. It indicates the set of SR nodes.
- **Segment:** An instruction that a node executes on an ingress packet (for example, forwarding a packet to a destination according to a shortest path, forwarding a packet by using a designated interface, or forwarding a packet to a specified application / service instance).
- **SID:** Segment ID. Segment Routing defines the destination address prefix / node and adjacency on the network as a segment and assigns segment IDs to these destination address prefixes / nodes and adjacencies. The segment ID is equivalent to the MPLS label in the traditional MPLS technology, and is mapped to the MPLS label at the forwarding plane.
- **SRGB:** Segment routing global block, Local tag set reserved for a global segment. In MPLS, SRGB is a set of local labels reserved for global labels. In IPv6, SRGB is a set of IPv6 addresses that have not been used globally.
- **Segment List:** It is a sort set of destination address prefix SID / node SID and adjacent SID ordered list, and is used to identify a complete label switching path LSP(Label Switched Path). In the MPLS architecture, the segment list is encapsulated in the packet header to guide forwarding.



## Segment

- The basic concept of SR is to divide the network into different Segments and then combine them to guide packet forwarding based on the specified path.
- SID: Segment ID, which identifies a unique segment. On the forwarding plane, MPLS labels are mapped.
- Three basic segments of SR: Adjacent segment, forward drop segment, and node segment.

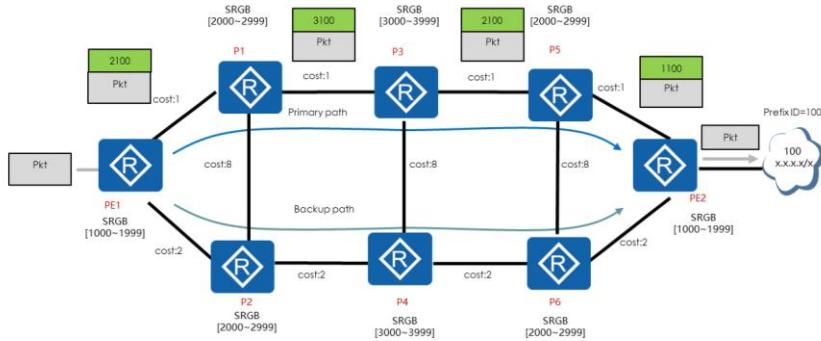
Segment	Generation Mode	Function	Example
Adjacency Segment	Dynamically allocated by source nodes through protocols	<p>It identifies an adjacency of a node on the network.</p> <p>It is allocated by the IGP protocol of the source node. It is valid only on the local node and has certain directivity.</p> <p>It is flooded to other NEs through IGP. It is globally visible and valid locally.</p> <p>It is identified by Adjacency Segment ID. Adjacency SID is a local SID outside the SRGB range.</p>	
Prefix Segment	Manually configured	<p>It identifies the prefix of a destination address on the network.</p> <p>It is globally visible and globally valid, and is flooded to other NEs through IGP.</p> <p>Prefix SID is the offset value in the SRGB range advertised by the source end. The receive end calculates the actual label value based on the SRGB value to generate MPLS forwarding entries.</p>	
Node Segment	Manually configured	It is a special Prefix Segment used to identify a specific node. Configure an IP address as the prefix of the loopback interface on the node. The Prefix SID of the node is Node SID.	

- Prefix Segment indicates the destination address, and Adjacency Segment indicates the outgoing link of the data packet, which can be similar to the destination IP address and outbound interface in the traditional IP forwarding. In an IGP area, the network element device floods the Node SID and the Adjacency SID of the network element device by using an extended IGP message, so that any network element can obtain information of another network element. Any path on the network can be constructed by combining the prefix (node) SID and adjacent SID in sequence. In each hop of the path, the next hop is distinguished by using the stack top segment information. The segment information is stacked on the top of the data head in sequence. When the stack top segment information includes the identifier of another node, the receiving node uses the equivalent multipath (ECMP) to forward the data packet to the next hop. When the stack top segment information is the identifier of the current node, the receiving node pops up the top segment and executes the task required by the next segment.
- In actual applications, Adjacency Segment, Prefix Segment, and Node Segment can be used independently or together.
- There are three scenarios: Prefix Segment, Adjacency Segment, Adjacency Segment + Node Segment



## Segment—Prefix Segment

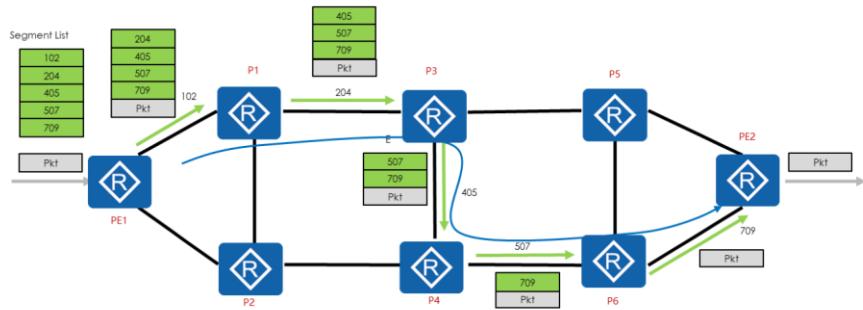
- Based on Prefix Segment: IGP uses the SPF algorithm to calculate the shortest path, which is also called SR-BE. As shown in the following figure, node Z is the destination node and its Prefix SID is 100. After IGP flooding is enabled, all devices in the IGP area learn the Prefix SID of PE2, and then use the SPF algorithm to obtain the shortest path to PE2.





## Segment —Adjacency Segment

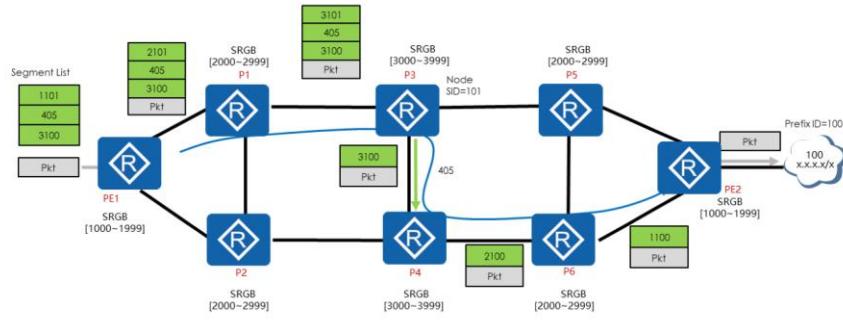
- Based on Adjacency Segment: The head node specifies a strict explicit path (Strict Explicit). In this manner, path adjustment and traffic adjustment can be performed in a centralized manner, so that Software-Defined Networking (SDN) can be better implemented. The Adjacency Segment is mainly used for the Traffic Engineering (SR-TE).





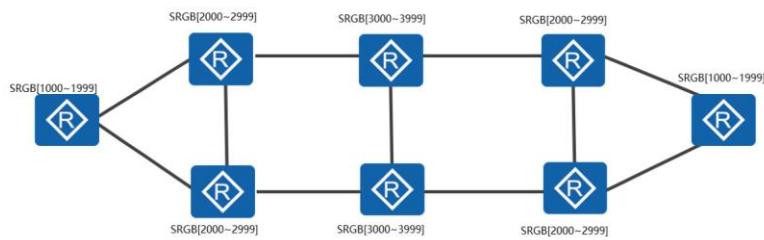
## Segment —Adjacency Segment + Node Segment

- Adjacency Segment +Node Segment: A combination of an explicit path and a shortest path is called a Loose Explicit (Loose Explicit). It is mainly used for SR-TE.





## SRGB



- The SRGB is a segment isolated from local label resources, and is specially used for Segment Routing, so that the Segment Routing global label and the traditional MPLS coexist locally.
- The node ID is the basis of the node SID. The globally unique node label must not conflict with the local label.
- The SRGB range and start value need to be configured. Based on the start value, the index value of the node SID is offset to obtain the local label.
- Why is the SRGB range of each device not the same?
  - The scope of SRGB is not specified in the standard, and the implementation of different vendors is different.
  - In the case of coexistence with traditional MPLS, the idle label space on each device cannot be the same.



# Contents

1. SR Technical Background

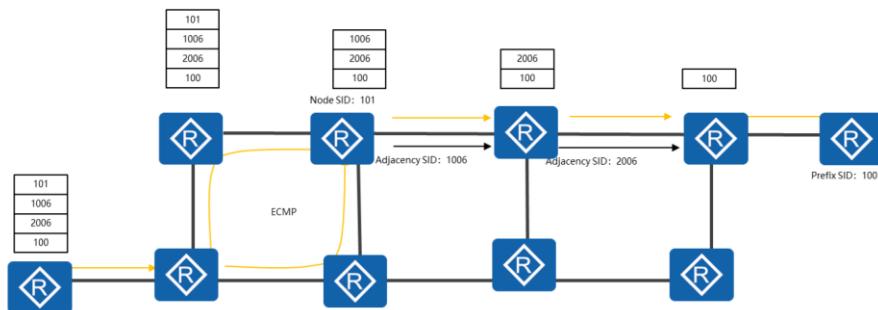
2. SR Principle Analysis

    2.1 Basic Concepts of SR

    2.2 SR Working Principle



## Principle of SR

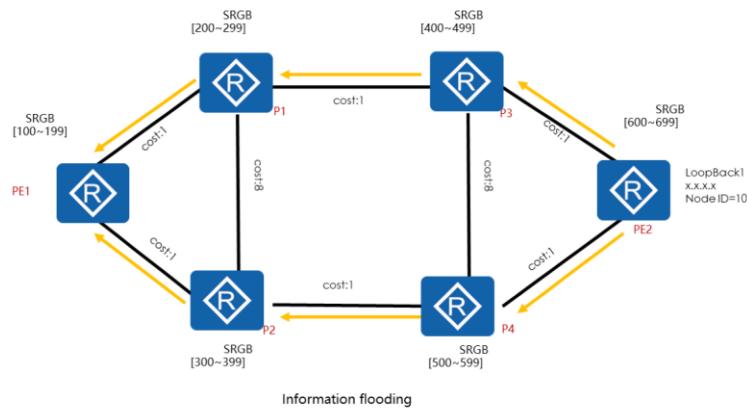


- SR is a protocol designed for forwarding data packets on the network based on the MPLS protocol and based on the source routing technology.
- SR divides the destination address prefix / node and adjacency in the network into segments and allocates SID (Segment ID) to these segments. The Adjacency SID (adjacent segment) and Prefix/Node SID (destination address prefix / node segment) are arranged in order to obtain a forwarding path.

- Segment Routing encapsulates the segment sequence representing the forwarding path into a data packet header and transmits the packet with the data packet. After receiving the data packet, the network node parses the segment sequence. If the top segment identifier of the segment sequence is Node SID, the network node forwards the segment sequence to the node according to the shortest path calculated by the SPF (if there is an equivalent path, the ECMP can be implemented), and if the path is a Adjacency SID, forwards the segment sequence to the next node according to the Adjacency SID, the data packet reaches the destination node.

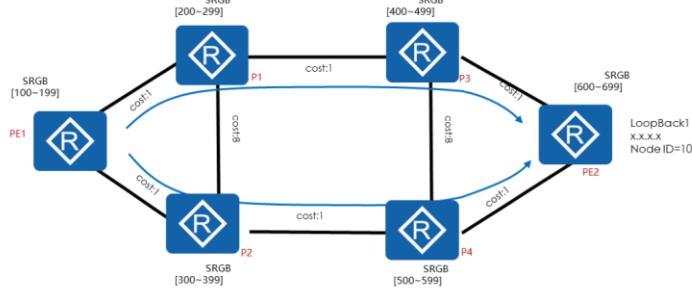


## Working Principle of SR- (IS-IS SR-BE) (Control Plane)





# Working Principle of SR- (IS-IS SR-BE) (Control Plane)



PE1→P1→P3→PE2: Label forwarding entry generated by each node for the Node SID 10

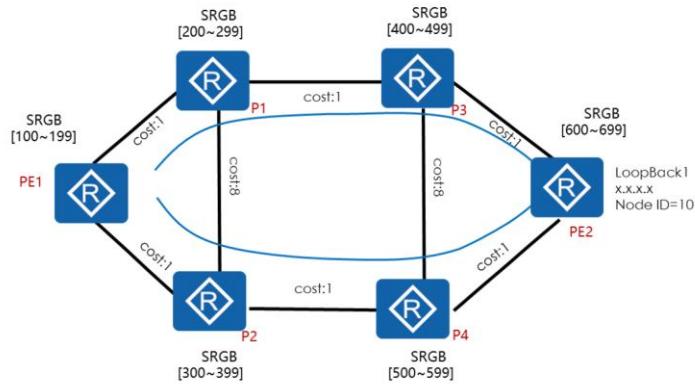
Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

PE1→P2→P4→PE2: Label forwarding entry generated by each node for the Node SID 10

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4->PE2
PE2	610	NA	NA

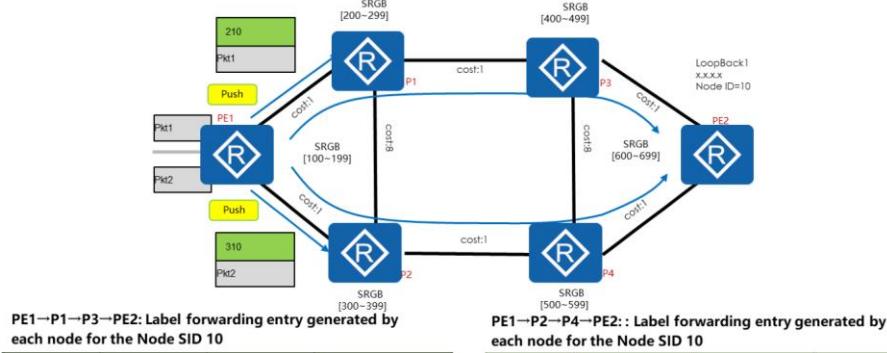


## Working Principle of SR- (IS-IS SR-BE) (Control Plane)





# Working Principle of SR- (IS-IS SR-BE) (Data Plane)



PE1→P1→P3→PE2: Label forwarding entry generated by each node for the Node SID 10

PE1→P2→P4→PE2: Label forwarding entry generated by each node for the Node SID 10

Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4->PE2
PE2	610	NA	NA

- The label operation type of the SR is the same as that of the MPLS, including label stack push, label stack swap, and label popping (Pop).

The following figure uses SR-BE Tunnel on PE1 to PE2 as an example.

1. PE1 on the ingress node: After receiving a service packet, the encapsulates the outgoing label allocated by the next hop P according to the destination address, label forwarding table, and ECMP traffic distribution algorithm, and then forwards the packet through the corresponding interface based on the destination prefix SID and label forwarding table.

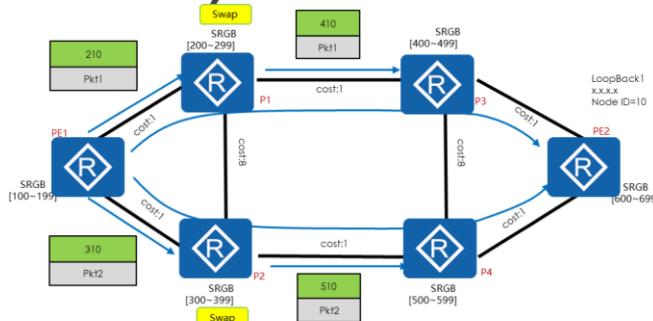
Example:

Push 210 to P1

Push 310 to P2



# Working Principle of SR- (IS-IS SR-BE) (Data Plane)



**PE1→P1→P3→PE2:** Label forwarding entry generated by each node for the Node SID 10

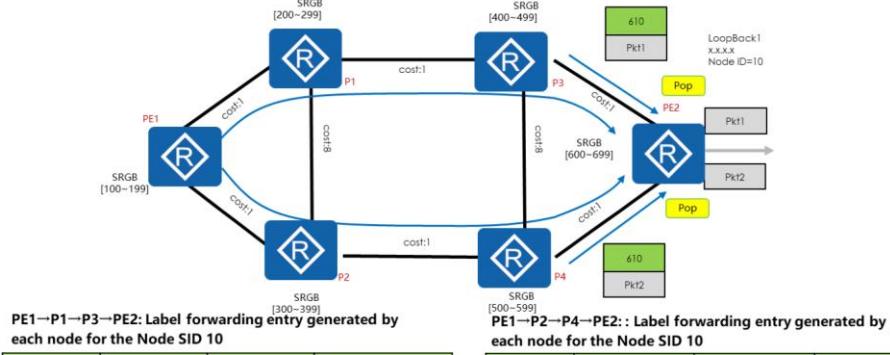
Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

**PE1→P2→P4→PE2:** Label forwarding entry generated by each node for the Node SID 10

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4->PE2
PE2	610	NA	NA



# Working Principle of SR- (IS-IS SR-BE) (Data Plane)

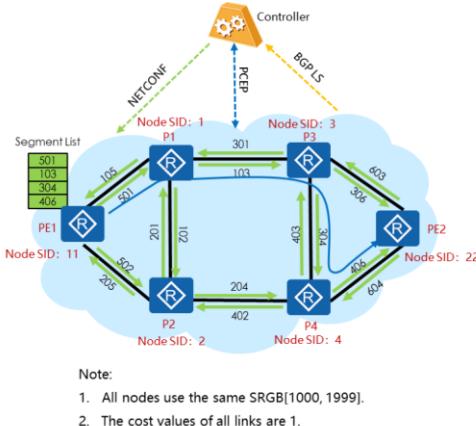


Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4->PE2
PE2	610	NA	NA



# Working Principle of SR- (IS-IS SR-TE) (Control Plane)



Label table on the controller

Node	Adjacency	label
PE1	NA	1011
PE1	Link PE1 to P1	501
PE1	Link PE1 to P2	502
P1	NA	1001
P1	Link P1 to PE1	105
P1	Link P1 to P2	102
P1	Link P1 to P3	103
.....	.....	.....

## 1. Generation of labels, label forwarding tables, and topology information

Each node on the network uses the IS-IS SR protocol to allocate adjacent SIDs to its adjacencies, floods the IDs to the entire network, and generates label forwarding entries for the local adjacent SIDs allocated to the nodes. In addition, the topology information with the adjacent SID information is generated.

In addition, the manually configured SRGB/prefix SID/node SID, SR capability, and SR algorithm are flooded in the IGP domain through IGP packets. Each node runs IS-IS SPF to calculate the shortest forwarding path of each node label, and generates a label forwarding table.

## 2. Report the label and topology information.

Each node in the network reports the topology information with the adjacent SID information to the controller through the BGP Link-State (BGP-LS).

## 3. Path calculation

The controller uses the Path Computation Element (PCEP) to calculate the label forwarding path.

## 4. Deliver trails.

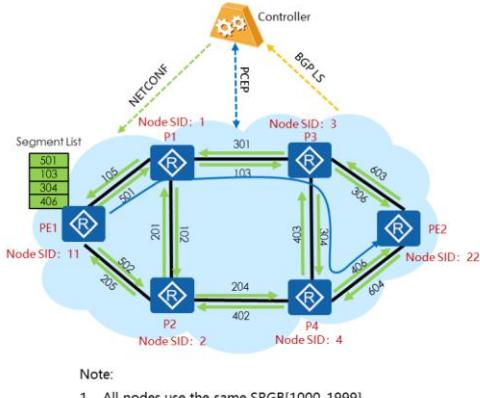
The controller delivers tunnel information through PCEP and delivers tunnel attributes to the first node of the tunnel through NETCONF. The first node of the tunnel reports the tunnel status to the controller through PCEP.

## 5. Create a tunnel.

The first node of the tunnel establishes a SR-TE Tunnel based on the label stack delivered by the controller.



# Working Principle of SR- (IS-IS SR-TE) (Control Plane)



Label forwarding entries on PE1

SID/Inlabel	Outlabel	Interface
1011	NA	NA
501	NA	PE1->P1
502	NA	PE1->P2
1001	1001	PE1->P1
1002	1002	PE1->P2
1003	1003	PE1->P1
1004	1004	PE1->P2
1022	1022	PE1->P1
1022	1022	PE1->P2

SR-TE Tunnel label stack encapsulation table on PE1

SR-TE Tunnel	Segment List
Tunnel 1	501/103/304/406
.....	.....

## 1. Generation of labels, label forwarding tables, and topology information

Each node on the network uses the IS-IS SR protocol to allocate adjacent SIDs to its adjacencies, floods the IDs to the entire network, and generates label forwarding entries for the local adjacent SIDs allocated to the nodes. In addition, the topology information with the adjacent SID information is generated.

In addition, the manually configured SRGB/prefix SID/node SID, SR capability, and SR algorithm are flooded in the IGP domain through IGP packets. Each node runs IS-IS SPF to calculate the shortest forwarding path of each node label, and generates a label forwarding table.

## 2. Report the label and topology information.

Each node in the network reports the topology information with the adjacent SID information to the controller through the BGP Link-State (BGP-LS).

## 3. Path calculation

The controller uses the Path Computation Element (PCEP) to calculate the label forwarding path.

## 4. Deliver trails.

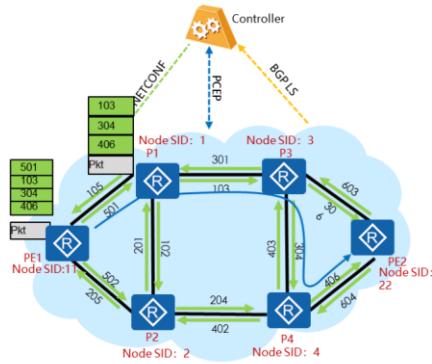
The controller delivers tunnel information through PCEP and delivers tunnel attributes to the first node of the tunnel through NETCONF. The first node of the tunnel reports the tunnel status to the controller through PCEP.

## 5. Create a tunnel.

The first node of the tunnel establishes a SR-TE Tunnel based on the label stack delivered by the controller.



# Working Principle of SR- (SR-TE Strict Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.

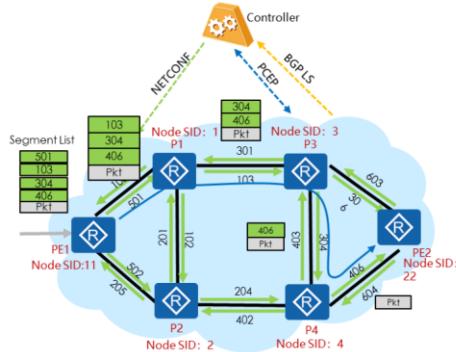
- SR node performs a label operation on the packet according to the label stack corresponding to the SR-TE tunnel in the packet header. Searches for the outbound interface hop by hop based on the top label of the stack to instruct data packets to be forwarded to the destination address of the tunnel.

Take the strict path SR-TE Tunnel from PE1 to PE2 as an example.

- 1. PE1 on the ingress node: After receiving the service packet, the encapsulates the label stack corresponding to the SR-TE Tunnel according to the SR-TE Tunnel determined by the routing policy or tunnel policy configured in the service. If it is determined that the corresponding stack top label 501 is an adjacent SID, the stack top label 501 is popped up, and the remaining part [103, 304, 406] of the label stack is encapsulated into the service packet, and is forwarded out from the corresponding interface PE1->P1 according to the top label and the label forwarding table.



# Working Principle of SR- (SR-TE Strict Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

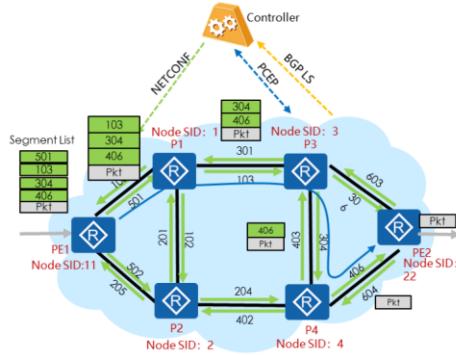
Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



# Working Principle of SR- (SR-TE Strict Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

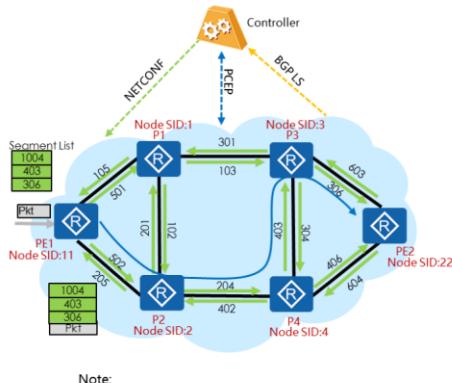
Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



# Working Principle of SR- (SR-TE Loose Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.

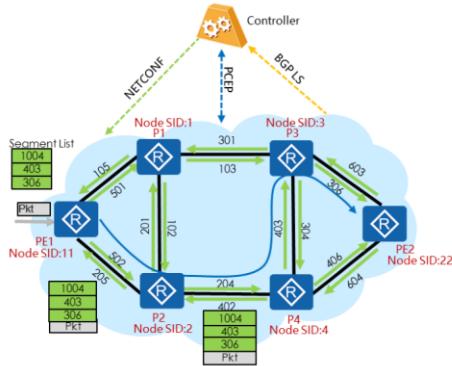
The SR performs a label operation on the packet according to the label stack corresponding to the SR-TE tunnel in the packet header. Searches for the outbound interface hop by hop based on the top label of the stack to instruct data packets to be forwarded to the destination address of the tunnel.

Take the loose path SR-TE Tunnel from PE1 to PE2 as an example.

- 1. PE1 on the ingress node: After receiving the service packet, the encapsulates the label stack [1004, 403, 306] corresponding to the SR-TE Tunnel according to the SR-TE Tunnel determined by the routing policy or tunnel policy configured in the service. Encapsulating the SR outgoing label 1004 allocated by the next hop P node according to the outer label and the label forwarding table, and forwarding the SR outgoing label 1004 according to the label forwarding table.



## Working Principle of SR- (SR-TE Loose Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

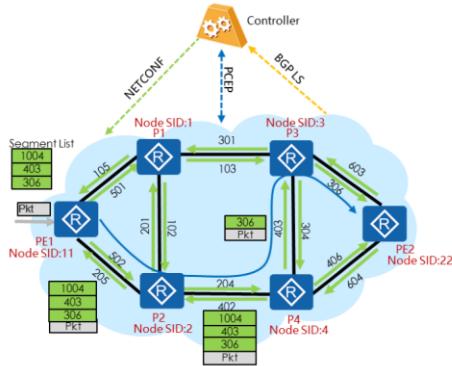
Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



# Working Principle of SR- (SR-TE Loose Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

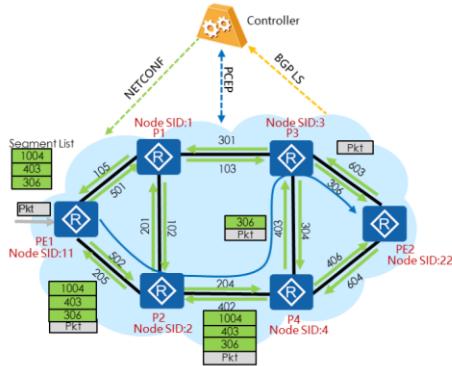
Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



# Working Principle of SR- (SR-TE Loose Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

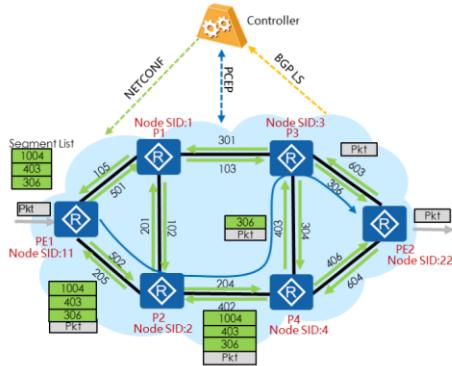
Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



# Working Principle of SR- (SR-TE Loose Explicit Path) (Data Plane)



The following table shows the label forwarding entries on each node on the SR-TE Tunnel path from PE1 to PE2.

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

Note:

1. All nodes use the same SRGB[1000, 1999].
2. The cost values of all links are 1.



## Quiz

1. Which of the followings are the SR tunnel type ?( )
  - A. SR-BE tunnel
  - B. SR-TE Loose Explicit Path tunnel
  - C. SR-BE Loose Explicit Path tunnel
  - D. SR-TE SR-TE Strict Explicit Path
2. Which of the following actions may be performed when packets are forwarded based on the SR label? ( )
  - A. Push
  - B. Pop
  - C. Swap
  - D. Switch

- Answer
- 1:ABD
- 2:ABC



## Summary

- SR Background
- Basic Concepts of SR
- SR technical architecture
- Working Principles of the SR

A blue-toned silhouette of a group of business people standing in a modern office environment with large windows and a grid pattern. The silhouettes are dark blue against a lighter blue background.

Thank You  
[www.huawei.com](http://www.huawei.com)



# Recommendations

- Huawei Learning Website
  - <http://learning.huawei.com/en>
- Huawei e-Learning
  - <https://ilearningx.huawei.com/portal/#/portal/EBG/51>
- Huawei Certification
  - [http://support.huawei.com/learning/NavigationAction!createNavi?navId=\\_31&lang=en](http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=en)
- Find Training
  - [http://support.huawei.com/learning/NavigationAction!createNavi?navId=\\_trainingsearch&lang=en](http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=en)



## More Information

- Huawei learning APP

