Basic Enterprise Network Architectures

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

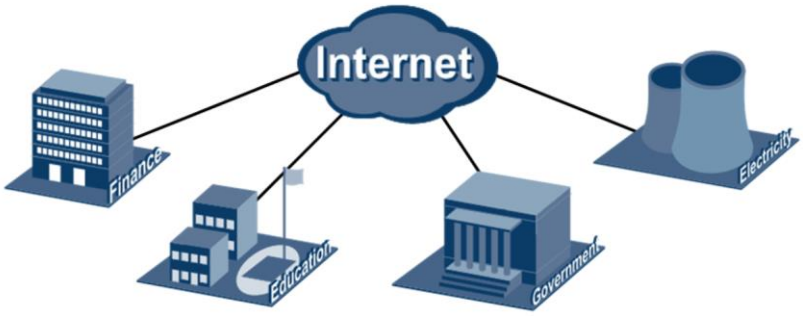
Enterprise business requirements highlight a need for networks that are capable of adapting to ever changing business demands in terms of enterprise business growth and evolving services. It is imperative therefore to understand the principles of what constitutes an enterprise network and how it is formed and adapted to support real world business demands.

Objectives

Upon completion of this section, trainees will be able to:

- Explain what constitutes an enterprise network
- Describe the common enterprise network architecture types
- Describe some of the solutions commonly implemented within an enterprise network to support business operations.

Real World Enterprise Networks



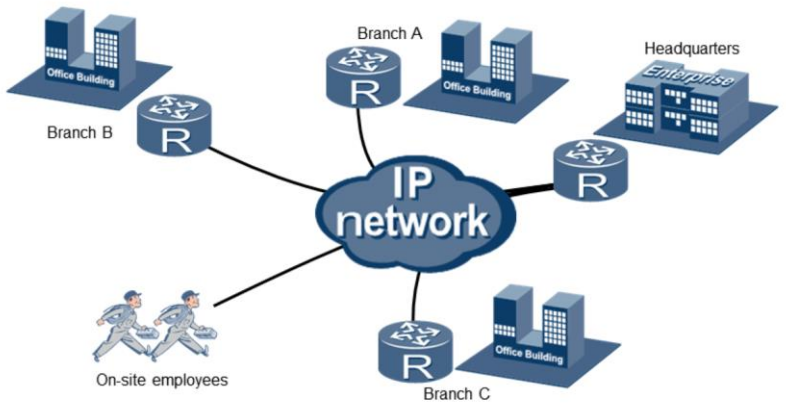
- Enterprise networks exist in many well known industries.
- Networks range from offices to large industrial platforms.

HUAWEI

The enterprise network originally represents the interconnection of systems belonging to a given functional group or organization to primarily enable the sharing of resources such as printers and file servers, communication support through means such as email, and the evolution towards applications that enable collaboration between users. Enterprise networks can be found today present within various industries from office environments to larger energy, finance and government based industries, which often comprise of enterprise networks that span multiple physical locations.

The introduction of the Internet as a public network domain allowed for an extension of the existing enterprise network to occur, through which geographically dispersed networks belonging to a single organization or entity could be connected, bringing with it a set of new challenges to establish interconnectivity between geographically dispersed enterprise networks, whilst maintaining the privacy and security of data belonging to an individual enterprise.

Enterprise Remote Networks

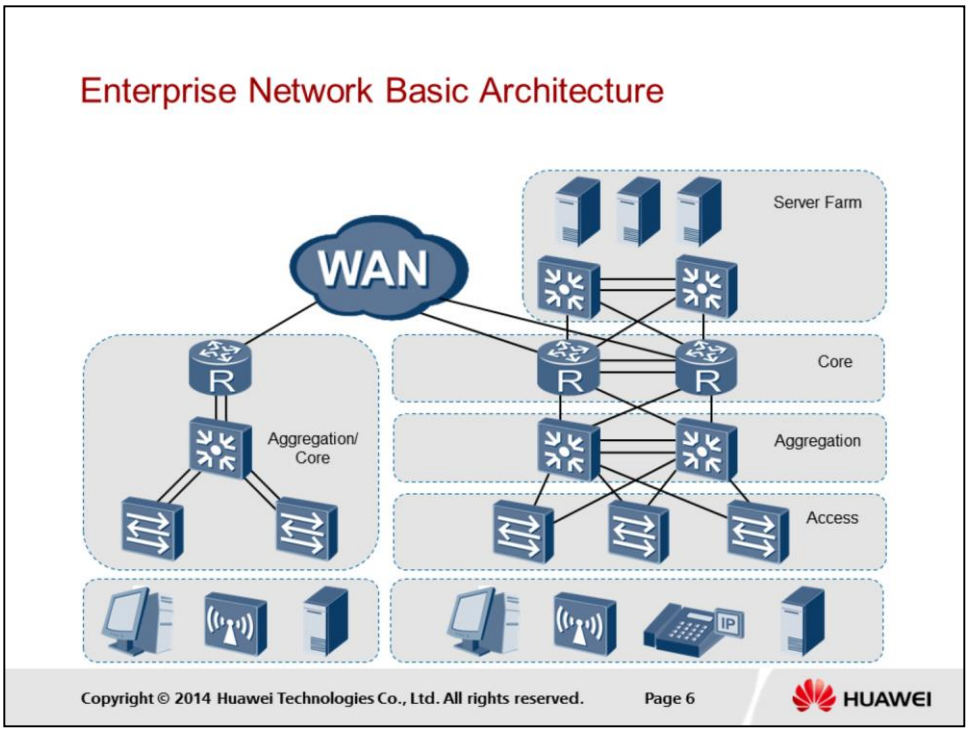


- Enterprise networks may span over large geographical distances.

HUAWEI

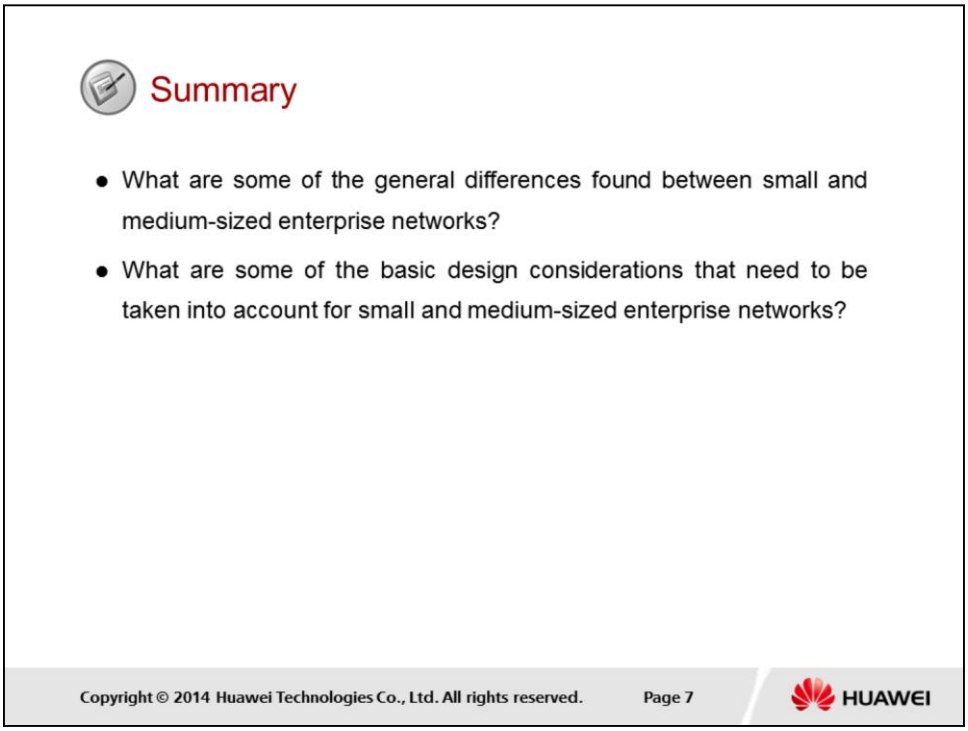
Various challenges impact today's industries in providing solutions for establishment of interconnectivity between remote locations, which often take the form of regional branch and head offices, as well as employees that represent a non fixed entity within the enterprise network, often being present in locations beyond the conventional boundaries of the existing enterprise. Challenges for industries have created a demand for ubiquitous networks that allow the enterprise network to be available from any location and at any time, to ensure access to resources and tools that allow for the effective delivery of support and services to industry partners and customers.

The evolution in enterprise solutions has enabled for public and third party IP networks to provide this anywhere anytime connectivity, along with the development of technologies that establish private network connections over this public network infrastructure, to extend the remote capabilities of the enterprise network beyond the physical boundaries of the enterprise, allowing remote office and users alike to establish a single enterprise domain that spans over a large geographic expanse.



Enterprise network architecture solutions vary significantly depending on the requirement of the industry and the organization. Smaller enterprise businesses may often have a very limited requirement in terms of complexity and demand, opting to implement a flat form of network, mainly due to the size of the organization that is often restricted to a single geographical location or within a few sites, supporting access to common resources, while enabling flexibility within the organization to support a smaller number of users. The cost to implement and maintain such networks is significantly reduced, however the network is often susceptible to failure due to lack of redundancy, and performance may vary based on daily operations and network demand.

Larger enterprise networks implement solutions to ensure minimal network failure, controlled access and provision for a variety of services to support the day-to-day operations of the organization. A multi layered architecture is defined to optimize traffic flow, apply policies for traffic management and controlled access to resources, as well as maintain network availability and stable operation through effective network redundancy. The multi layer design also enables easy expansion, and together with a modular design that provides for effective isolation and maintenance should problems in the network occur, without impacting the entire network.



1. Small enterprise networks that implement a flat network architecture may limit the capability to scale the network in the event of growth in the number of users. Where it is expected that a larger number of users will need to be supported, a hierarchical approach to enterprise networks should be considered. Medium-sized networks will generally support a greater number of users, and therefore will typically implement a hierarchical network infrastructure to allow the network to grow and support the required user base.

2. Small and medium sized enterprise networks must take into account the performance of the network as well as providing redundancy in the event of network failure in order to maintain service availability to all users. As the network grows, the threat to the security of the network also increases which may also hinder services.

Thank you

www.huawei.com

Enterprise Network Constructs

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

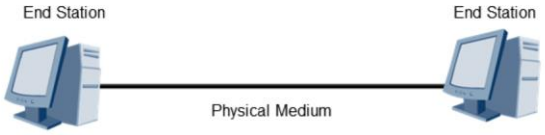
Establishment of an enterprise network requires a fundamental understanding of general networking concepts. These concepts include knowledge of what defines a network, as well as the general standards of technology and physical components that are used to establish enterprise networks. An understanding of the underlying network communications and the impact that such behavior has on the network is also paramount to ensuring performance effective implementation.

Objectives

Upon completion of this section, you will be able to:

- Explain what constitutes a network.
- Identify the basic components of a network.
- Describe the primary mechanisms for communication over a network.

Simple Point to Point Ethernet Networks



- Networks are comprised of at least two end stations, and a medium over which data can be carried.

HUAWEI

A network can be understood to be the capability of two or more entities to communicate over a given medium. The development of any network relies on this same principle for establishing communication. Commonly the entities within a network that are responsible for the transmission and reception of communication are known as end stations, while the means by which communication is enabled is understood to be the medium. Within an enterprise network, the medium exists in a variety of forms from a physical cable to radio waves.

Coaxial

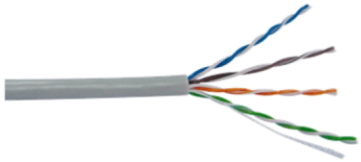
Standard	Cables	Maximum Transmission Distance
10Base2	Thin coaxial	185m
10Base5	Thick coaxial	500m

- Copper coaxial cabling commonly used to support users as part of a shared network.

The coaxial cable represents a more historic form of transmission medium that may today be limited in usage within the enterprise network. As a transmission medium, the coaxial cable comprises generally of two standards, the 10Base2 and 10Base5 forms, that are known as Thinnet or Thinwire, and Thicknet or Thickwire respectively.

The standards both support a transmission capacity of 10Mbps transmitted as baseband signals for respective distances of 185 and 500 meters. In today's enterprise networks, the transmission capacity is extremely limited to be of any significant application. The Bayonet Neill-Concelman (BNC) connector is the common form of connector used for thin 10Base2 coaxial cables, while a type N connector was applied to the thicker 10Base5 transmission medium.

Ethernet



Standard	Physical Medium	Distance
10Base-T	Two pairs of Category 3/4/5 twisted pair cables	100m
100Base-TX	Two pairs of Category 5 twisted pair cables	100m
1000Base-T	Four pairs of Category 5e twisted pair cables	100m

- The primary physical medium used in enterprise networks.

Ethernet cabling has become the standard for many enterprise networks providing a transmission medium that supports a much higher transmission capacity. The medium supports a four copper wire pair contained within a sheath which may or may not be shielded against external electrical interference. The transmission capacity is determined mainly based on the category of cable with category 5 (CAT5) supporting Fast Ethernet transmission capacity of up to 100Mbps, while a higher Gigabit Ethernet transmission capacity is supported from Category 5 extended (CAT5e) standards and higher.

The transmission over Ethernet as a physical medium is also susceptible to attenuation, causing the transmission range to be limited to 100 meters. The RJ-45 connector is used to provide connectivity with wire pair cabling requiring specific pin ordering within the RJ-45 connector, to ensure correct transmission and reception by end stations over the transmission medium.

Fiber Optic

Standard	Physical Medium	Distance
10Base-F	Two strand fiber	2000m
100Base-FX	Two strand multi-mode fiber	2000m
1000Base-LX	Single-mode fiber or multi-mode fiber	316 - 5000m
1000Base-SX	Multi-mode fiber	275 - 550m

Optical media uses light as a means of signal transmission as opposed to electrical signals found within both Ethernet and coaxial media types. The optical fiber medium supports a range of standards of 10Mbps, 100Mbps, 1Gbps and also 10Gbps (10GBASE) transmission. Single or multi-mode fiber defines the use of an optical transmission medium for propagating of light, where single mode refers to a single mode of optical transmission being propagated, and is used commonly for high speed transmission over long distances.

Multi mode supports propagation of multiple modes of optical transmission that are susceptible to attenuation as a result of dispersion of light along the optical medium, and therefore is not capable of supporting transmission over longer distances. This mode is often applied to local area networks which encompass a much smaller transmission range. There are an extensive number of fiber connector standards with some of the more common forms being recognized as the ST connector, LC connector and SC, or snap connector.

Serial

Standard	Speed
RS-232	Standards define up to 20000bps, but can reach 1Mbit/s
RS-422	100Kbit/s ~ 10Mbit/s+

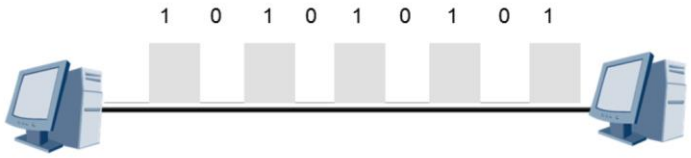
- Serial represents a legacy form of data transmission
- Standards continue to evolve as in forms such as USB.

HUAWEI

Serial represents a standard initially developed over 50 years ago to support reliable transmission between devices, during which time many evolutions of the standard have taken place. The serial connection is designed to support the transmission of data as a serial stream of bits. The common standard implemented is referred to as (Recommended Standard) RS-232 but it is limited somewhat by both distance and speed. Original RS-232 standards define that communication speeds supported be no greater that 20Kbps, based on a cable length of 50ft (15 meters), however transmission speeds for serial is unlikely to be lower than 115 Kbps. The general behavior for serial means that as the length of the cable increases, the supported bit rate will decrease, with an approximation that a cable of around 150 meters, or 10 times the original standards, the supported bit rate will be halved.

Other serial standards have the capability to achieve much greater transmission ranges, such as is the case with the RS-422 and RS-485 standards that span distances of up to 4900ft (1200 meters) and are often supported by V.35 connectors that were made obsolete during the late 1980's but are still often found and maintained today in support of technologies such as Frame Relay and ATM, where implemented. RS-232 itself does not define connector standards, however two common forms of connector that support the RS-232 standard include the DB-9 and DB-25 connectors. Newer serial standards have been developed to replace much of the existing RS-232 serial technology, including both FireWire and the universal serial bus (USB) standards, that latter of which is becoming common place in many newer products and devices.

Signal Data Encoding



0000	0001	0010	0011	0100
0	1	2	3	4

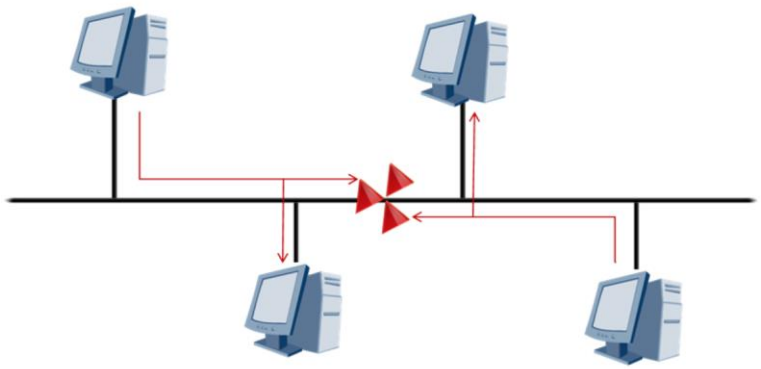
- Signal patterns used for interpretation of communication.
- Encoding is used to synchronize transmission.

In order to enable communication over physical links, signals must be transmitted between the transmitting and receiving stations. This signal will vary depending on the medium that is being used, as in the case of optical and wireless transmission. The main purpose of the signal is to ensure that synchronization (or clocking) between the sender and receiver over a physical medium is maintained, as well as support transmission of the data signal in a form that can be interpreted by both the sender and receiver.

A waveform is commonly recognized as a property of line encoding where the voltage is translated into a binary representation of 0 and 1 values that can be translated by the receiving station. Various line coding standards exist, with 10Base Ethernet standards supporting a line encoding standard known as Manchester encoding. Fast Ethernet with a frequency range of 100MHz invokes a higher frequency than can be supported when using Manchester encoding.

An alternative form of line encoding is therefore used known as NZRI, which in itself contains variations dependant on the physical media, thus supporting MLT-3 for 100Base-TX and 100Base-FX together with extended line encoding known as 4B/5B encoding to deal with potential clocking issues. 100Base-T4 for example uses another form known as 8B/6T extended line encoding. Gigabit Ethernet supports 8B/10B line encoding with the exception of 1000Base-T which relies on a complex block encoding referred to as 4D-PAM5.

Collision Domains



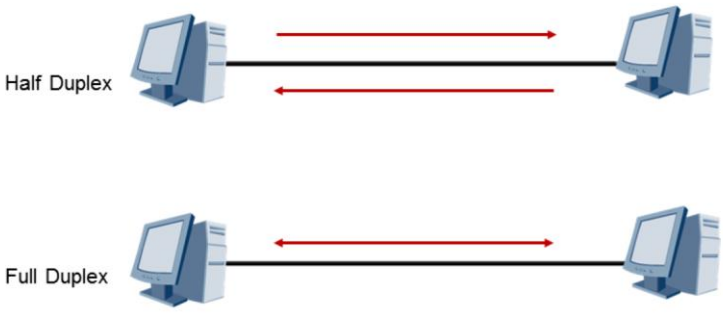
- Signals in a shared network are susceptible to collisions
- A collision detection mechanism is used to identify collisions.

HUAWEI

Ethernet represents what is understood to be a multi-access network, in which two or more end stations share a common transmission medium for the forwarding of data. The shared network is however susceptible to transmission collisions where data is forwarded by end stations simultaneously over a common medium. A segment where such occurrences are possible is referred to as a shared collision domain.

End stations within such a collision domain rely on contention for the transmission of data to an intended destination. This contentious behavior requires each end station monitor for incoming data on the segment before making any attempt to transmit, in a process referred to as Carrier Sense Multiple-Access Collision Detection (CSMA/CD). However, even after taking such precautions the potential for the occurrence of collisions as a result of simultaneous transmission by two end stations remains highly probable.

Duplex Modes



- Duplex modes support simultaneous and non-simultaneous bidirectional communication.

HUAWEI

Transmission modes are defined in the form of half and full duplex, to determine the behavior involved with the transmission of data over the physical medium.

Half duplex refers to the communication of two or more devices over a shared physical medium in which a collision domain exists, and with it CSMA/CD is required to detect for such collisions. This begins with the station listening for reception of traffic on its own interface, and where it is quiet for a given period, will proceed to transmit its data. If a collision were to occur, transmission would cease, followed by initiation of a backoff algorithm to prevent further transmissions until a random value timer expires, following which retransmission can be reattempted.

Full duplex defines the simultaneous bidirectional communication over dedicated point to point wire pairs, ensuring that there is no potential for collisions to occur, and thus there is no requirement for CSMA/CD.

Summary

- Which forms of cabling can be used to support Gigabit Ethernet transmissions within an enterprise network?
- What is a collision domain?
- What is the purpose of CSMA/CD?

1. Gigabit Ethernet transmission is supported by CAT 5e cabling and higher, and also any form of 1000Base Fiber Optic cabling or greater.
2. A collision domain is a network segment for which the same physical medium is used for bi-directional communication. Data simultaneously transmitted between hosts on the same shared network medium is susceptible to a collision of signals before those signals reach the intended destination. This generally results in malformed signals either larger or smaller than the acceptable size for transmission (64 bytes – 1500 bytes), also know as runts and giants, being received by the recipient.
3. CSMA/CD is a mechanism for detecting and minimizing the possibility of collision events that are likely to occur in a shared network. CSMA requires that the transmitting host first listen for signals on the shared medium prior to transmission. In the event that no transmissions are detected, transmission can proceed. In the unfortunate circumstance that signals are transmitted simultaneously and a collision occurs, collision detection processes are applied to cease transmission for a locally generated period of time, to allow collision events to clear and to avoid further collisions from occurring between transmitting hosts.

Thank you

www.huawei.com

Ethernet Framing

HUAWEI TECHNOLOGIES CO., LTD.

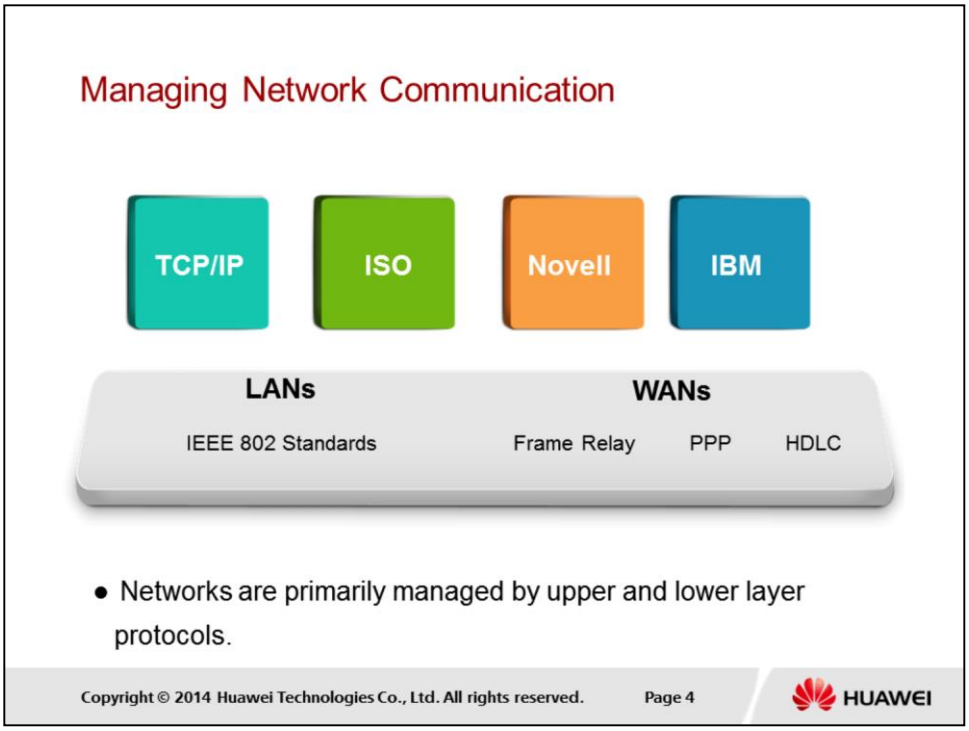
Foreword

Transmission over a physical medium requires rules that define the communication behavior. The management of the forwarding behavior of Ethernet based networks is controlled through IEEE 802 standards defined for Ethernet data link technology. A fundamental knowledge of these standards is imperative to fully understand how link layer communication is achieved within Ethernet based networks.

Objectives

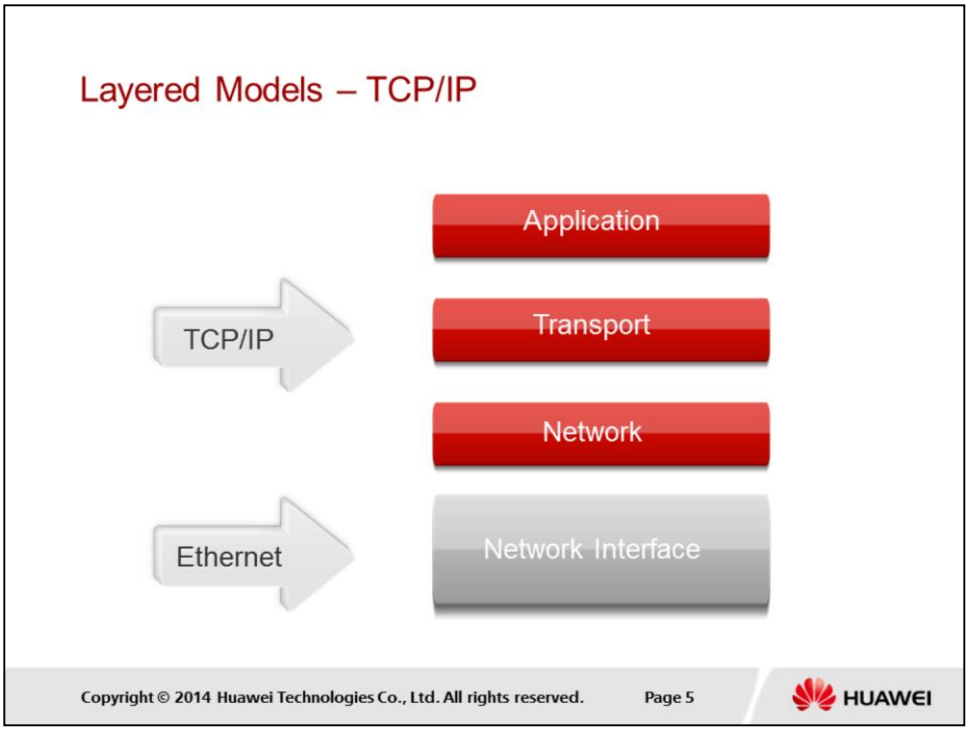
Upon completion of this section, trainees will be able to:

- Explain the application of reference models to networks.
- Describe how frames are constructed.
- Explain the function of MAC addressing at the data link layer.
- Describe Ethernet frame forwarding and processing behavior.



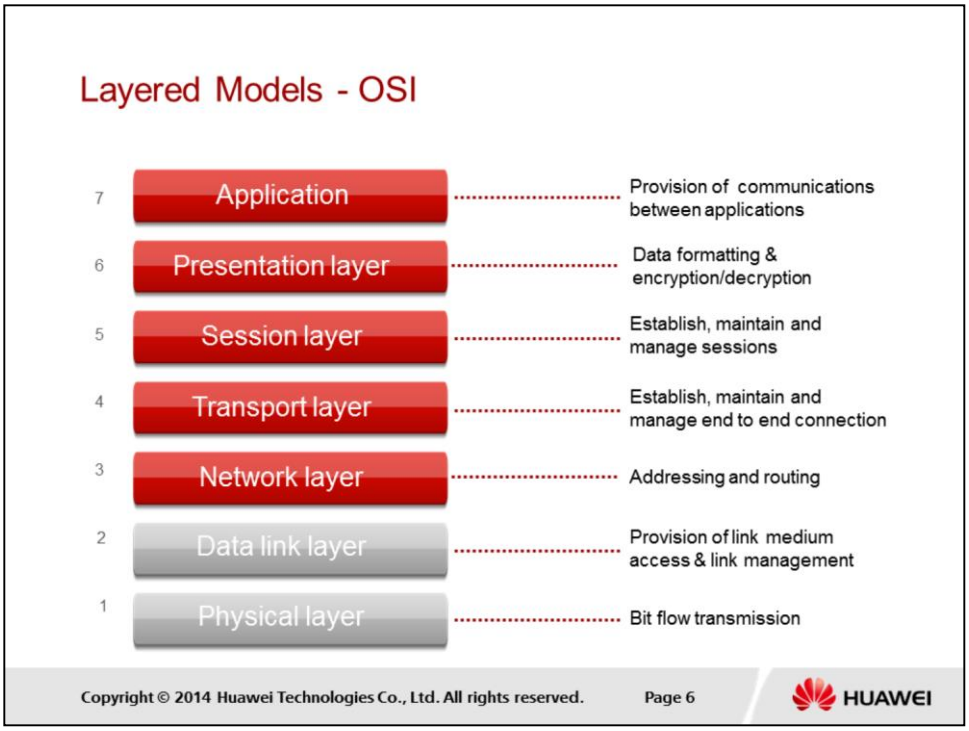
Communication over networks relies on the application of rules that govern how data is transmitted and processed in a manner that is understood by both the sending and receiving entities. As a result, multiple standards have been developed over the course of time with some standards becoming widely adopted. There exists however a clear distinction between the standards that manage physical data flow and the standards responsible for logical forwarding and delivery of traffic.

The IEEE 802 standards represent a universal standard for managing the physical transmission of data across the physical network and comprises of standards including the Ethernet standard 802.3 for physical transmission over local area networks. Alternative standards exist for transmission over wide area networks operating over serial based media, including Frame Relay, HDLC and more legacy standards such as ATM. TCP/IP has been widely adopted as the protocol suite defining the upper layer standards, regulating the rules (protocols) and behavior involved in managing the logical forwarding and delivery between end stations.



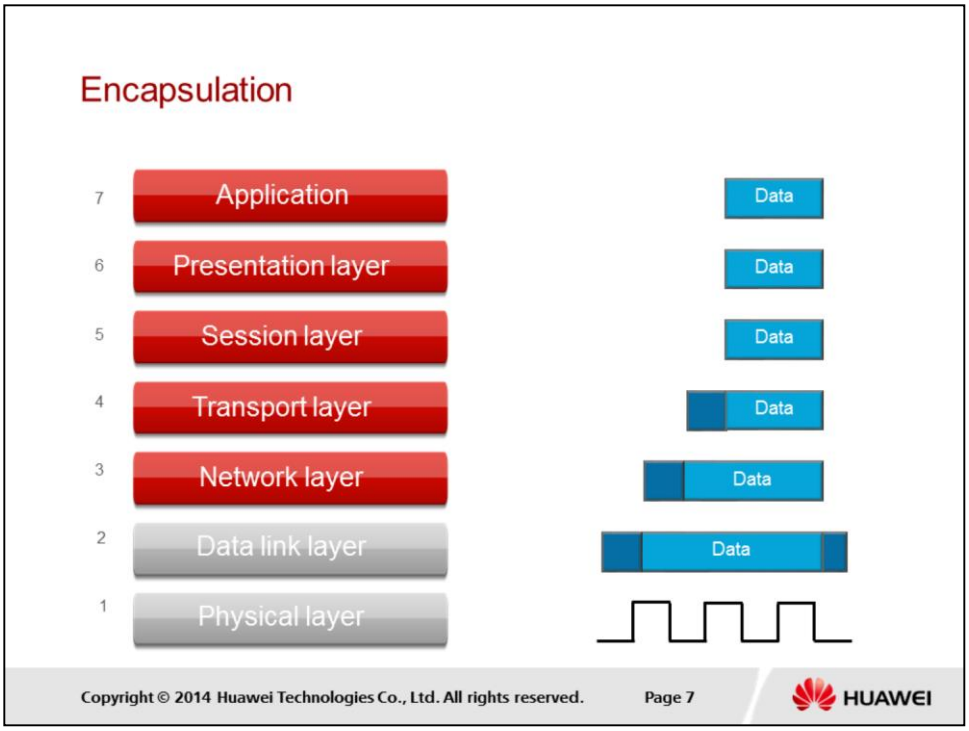
The TCP/IP reference model primarily concerns with the core principles of the protocol suite, which can be understood as the logical transmission and delivery of traffic between end stations. As such the TCP/IP protocol reference model provides a four layer representation of the network, summarizing physical forwarding behavior under the network interface layer, since lower layer operation is not the concern of the TCP/IP protocol suite.

Primary focus remains on the network (or Internet) layer which deals with how traffic is logically forwarded between networks, and the transport (sometimes referred to as host-to-host) layer that manages the end-to-end delivery of traffic, ensuring reliability of transportation between the source and destination end stations. The application layer represents an interface through a variety of protocols that enable services to be applied to end user application processes.



Although the TCP/IP reference model is primarily supported as the standard model based on TCP/IP protocol suite, the focus of the TCP/IP reference model does not clearly separate and distinguish the functionality when referring lower layer physical transmission.

In light of this, the open systems interconnection, or OSI reference model is often recognized as the model for reference to IEEE 802 standards due to the clear distinction and representation of the behavior of lower layers which closely matches the LAN/MAN reference model standards that are defined as part of the documented IEEE 802-1990 standards for local and metropolitan area networks. In addition the model, that is generally in reference to the ISO protocol suite, provides an extended breakdown of upper layer processing.



As upper layer application data is determined for transmission over a network from an end system, a series of processes and instructions must be applied to the data before transmission can be successfully achieved. This process of appending and pre-pending instructions to data is referred to as encapsulation and for which each layer of the reference model is designed to represent.

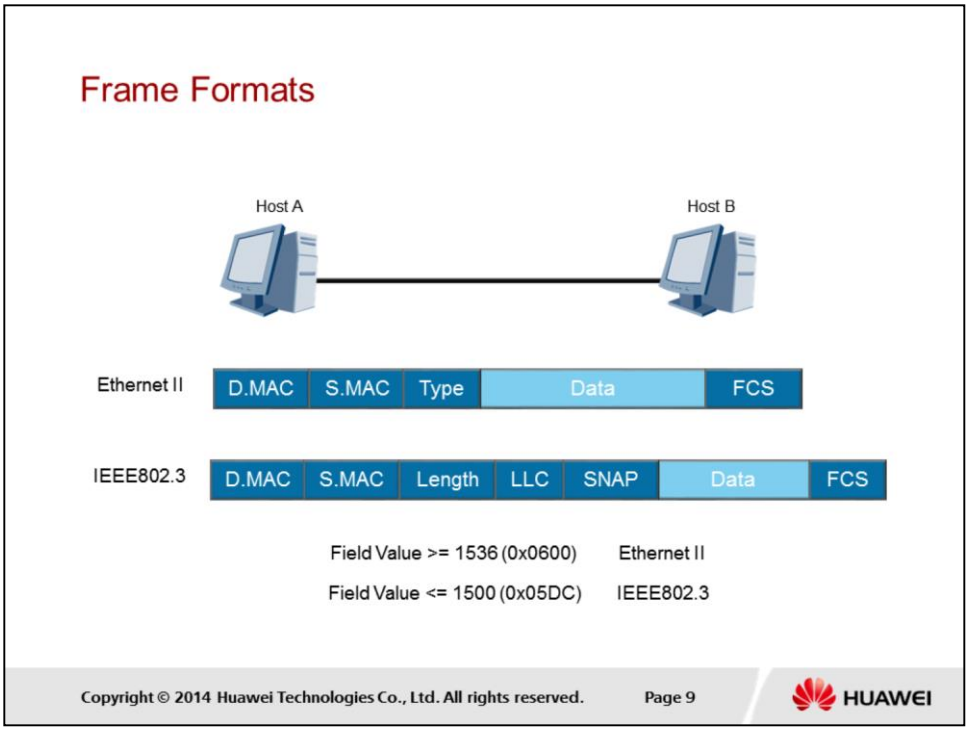
As instructions are applied to the data, the general size of the data increases. The additional instructions represent overhead to the existing data and are recognized as instructions to the layer at which the instructions were applied. To other layers, the encapsulated instructions are not distinguished from the original data. The final appending of instructions is performed as part of the lower layer protocol standards (such as the IEEE 802.3 Ethernet standard) before being carried as an encoded signal over a physical medium.

Communication Between Two End Stations

- Data link layer frames are used to govern transmission over the communications medium

As part of the IEEE 802.3 Ethernet standard, data is encapsulated with instructions in the form of a header and a trailer before it can be propagated over physical media on which Ethernet is supported. Each stage of encapsulation is referred to by a protocol data unit or PDU, which at the data link layer is known as a frame.

Ethernet frames contain instructions that govern how and whether data can be transmitted over the medium between two or more points. Ethernet frames come in two general formats, the selection of which is highly dependant on the protocols that have been defined prior to the framing encapsulation.

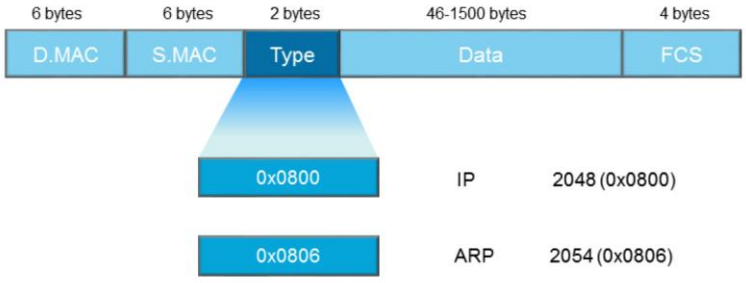


Two frame formats are recognized as standard for Ethernet based networks. The DIX version 2 frame type standard was originally developed during the early 1980's, where today it is recognized as the Ethernet II frame type. Ethernet II was eventually accepted and integrated into the IEEE 802 standards, highlighted as part of section 3.2.6 of the IEEE 802.3x-1997 standards documentation. The IEEE 802.3 Ethernet standard was originally developed in 1983, with key differences between the frame formats including a change to the type field that is designed to identify the protocol to which the data should be forwarded to once the frame instructions have been processed. In the IEEE 802.3 Ethernet format, this is represented as a length field which relies on an extended set of instructions referred to as 802.2 LLC to identify the forwarding protocol.

Ethernet II and IEEE 802.3 associate with upper layer protocols that are distinguished by a type value range, where protocols supporting a value less than or equal to 1500 (or 05DC in Hexadecimal) will employ the IEEE 802.3 Ethernet frame type at the data link layer. Protocols represented by a type value greater than or equal to 1536 (or 0600 in Hexadecimal) will employ the Ethernet II standard, and which represents the majority of all frames within Ethernet based networks.

Other fields found within the frame include the destination and source MAC address fields that identify the sender and the intended recipient(s), as well as the frame check sequence field that is used to confirm the integrity of the frame during transmission.

Ethernet II Frame

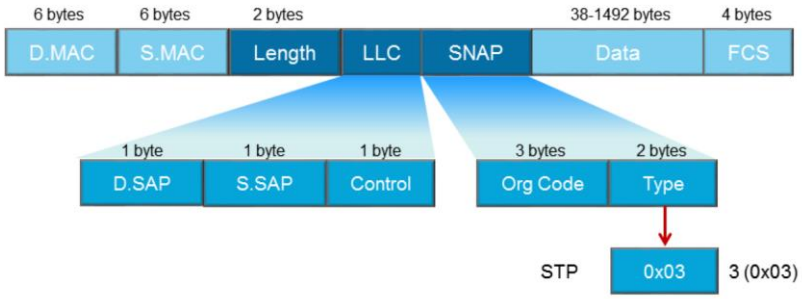


- The Ethernet II frame type is associated with protocols with a type value greater than 1536 (0x600).

HUAWEI

The Ethernet II frame references a hexadecimal type value which identifies the upper layer protocol. One common example of this is the Internet Protocol (IP) which is represented by a hexadecimal value of 0x0800. Since this value for IP represents a value greater than 0x0600 , it is determined that the Ethernet II frame type should be applied during encapsulation. Another common protocol that relies on the Ethernet II frame type at the data link layer is ARP, and is represented by the hexadecimal value of 0x0806.

IEEE802.3 Frame

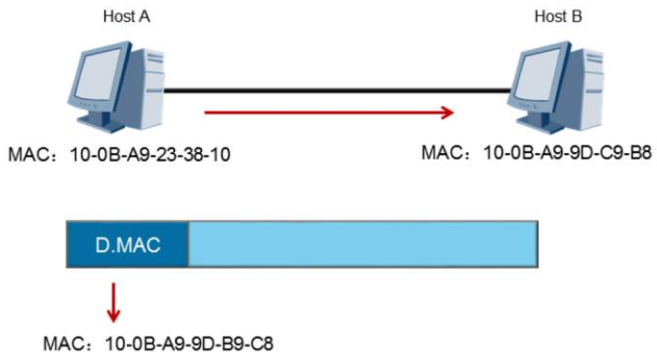


- The IEEE 802.3 frame type is associated with protocols with a type value less than 1500 (0x05DC).

HUAWEI

For the IEEE 802.3 frame type, the type field is contained as part of the SNAP extension header and is not so commonly applied the protocols in today's networks, partially due to the requirement for additional instructions which results in additional overhead per frame. Some older protocols that have existed for many years but that are still applied in support of Ethernet networks are likely to apply the IEEE 802.3 frame type. One clear example of this is found in the case of the Spanning Tree Protocol (STP) that is represented by a value of 0x03 within the type field of the SNAP header.

Frame Forwarding



- Media Access Control (MAC) addressing facilitates data link layer communication

HUAWEI

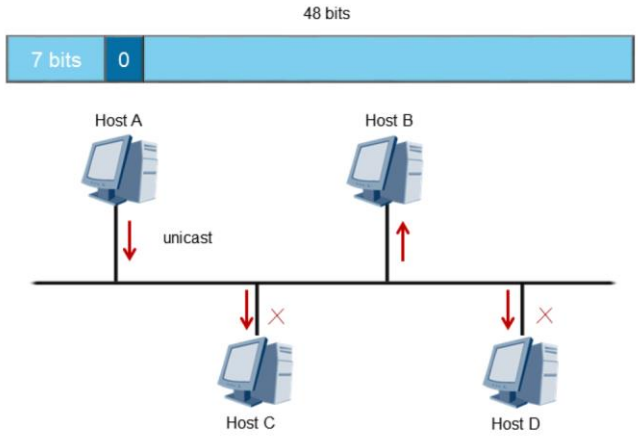
Ethernet based networks achieve communication between two end stations on a local area network using Media Access Control (MAC) addressing that allows end systems within a multi access network to be distinguished. The MAC address is a physical address that is burned into the network interface card to which the physical medium is connected. This same MAC address is retrieved and used as the destination MAC address of the intended receiver by the sender, before the frame is transferred to the physical layer for forwarding over the connected medium.

The Ethernet MAC Address

- MAC addresses are comprised of an organizationally unique identifier and a vendor assigned address value.

Each MAC address is a 48 bit value commonly represented in a hexadecimal (base 16) format and comprised of two parts that attempt to ensure that every MAC address is globally unique. This is achieved by the defining of an organizationally unique identifier that is vendor specific, based on which it is possible to trace the origin of a product back to its vendor based on the first 24 bits of the MAC address. The remaining 24 bits of the MAC address is a value that is incrementally and uniquely assigned to each product (e.g. a Network Interface Card or similar product supporting port interfaces for which a MAC is required).

Unicast Frame Forwarding

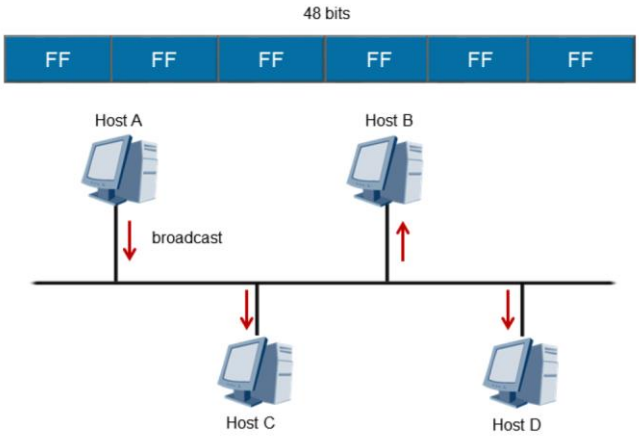


HUAWEI

The transmission of frames within a local network is achieved using one of three forwarding methods, the first of these is unicast and refers to the transmission from a single source location to a single destination. Each host interface is represented by a unique MAC address, containing an organizationally unique identifier, for which the 8th bit of the most significant octet (or first byte) in the MAC address field identifies the type of address. This 8th bit is always set to 0 where the MAC address is a host MAC address, and signifies that any frame containing this MAC address in the destination MAC address field is intended for a single destination only.

Where hosts exist within a shared collision domain, all connected hosts will receive the unicast transmission but the frame will be generally ignored by all hosts where the MAC address in the destination MAC field of the frame does not match the MAC value of the receiving host on a given interface, leaving only the intended host to accept and process the received data. Unicast transmissions are only forwarded from a single physical interface to the intended destination, even in cases where multiple interfaces may exist.

Broadcast Frame Forwarding

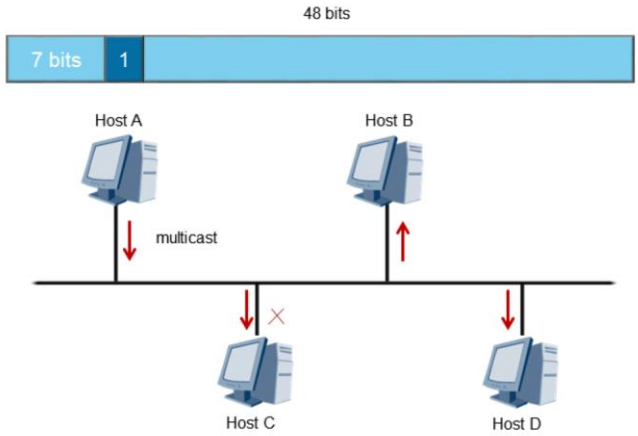


HUAWEI

Broadcast transmission represents a forwarding method that allows frames to be flooded from a single source received by all destinations within a local area network. In order to allow traffic to be broadcasted to all hosts within a local area network, the destination MAC address field of the frame is populated with a value that is defined in hexadecimal as FF:FF:FF:FF:FF:FF, and which specifies that all recipients of a frame with this address defined should accept receipt of this frame and process the frame header and trailer.

Broadcasts are used by protocols to facilitate a number of important network processes including discovery and maintenance of network operation, however also generate excessive traffic that often causes interrupts to end systems and utilization of bandwidth that tend to reduce the overall performance of the network.

Multicast Frame Forwarding

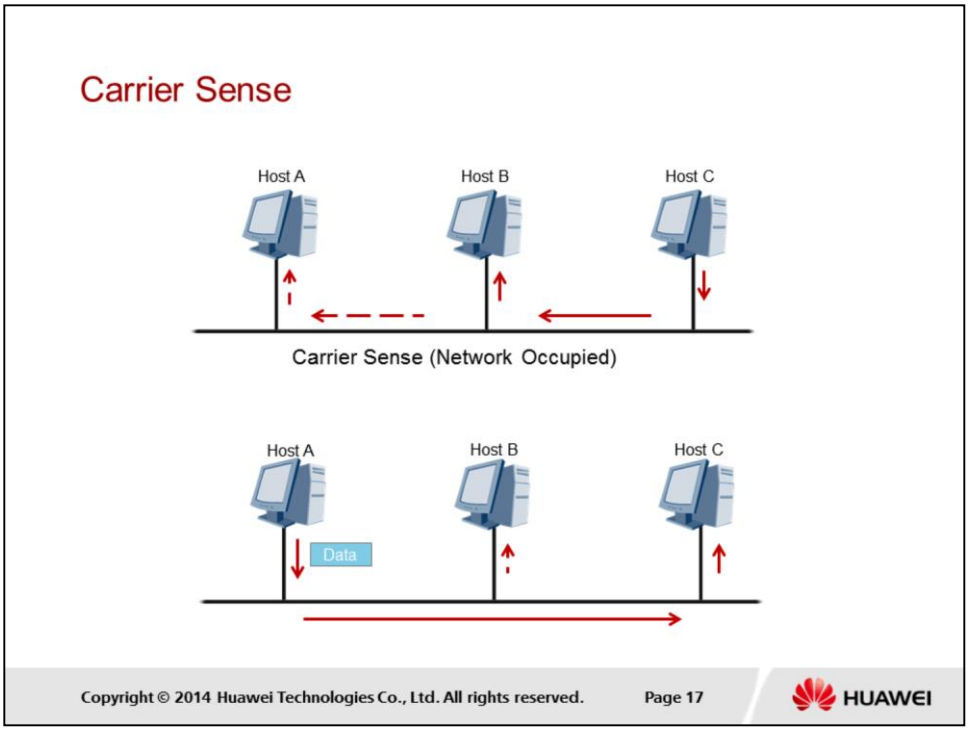


HUAWEI

A more efficient alternative to broadcast that has begun to replace the use of broadcasts in many newer technologies is the multicast frame type. Multicast forwarding can be understood as a form of selective broadcast that allows select hosts to listen for a specific multicast MAC address in addition to the unicast MAC address that is associated with the host, and process any frames containing the multicast MAC address in the destination MAC field of the frame.

Since there is no relative distinction between unicast MAC addresses and multicast MAC address formats, the multicast address is differentiated using the 8th bit of the first octet. Where this bit value represents a value of 1, it identifies that the address is part of the multicast MAC address range, as opposed to unicast MAC addresses where this value is always 0.

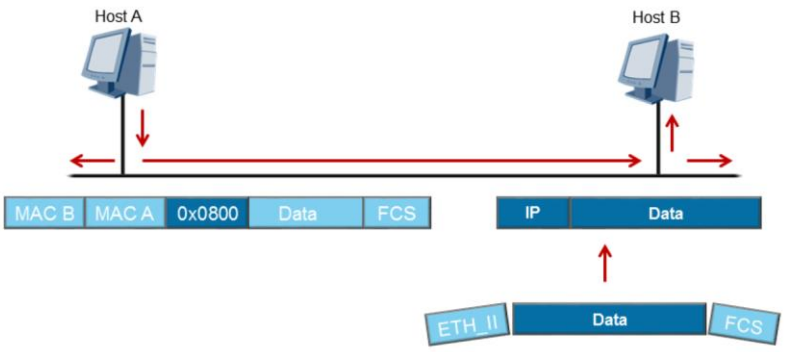
In a local area network, the true capability of multicast behavior at the data link layer is limited since forwarding remains similar to that of a broadcast frame in which interrupts are still prevalent throughout the network. The only clear difference with broadcast technology is in the selective processing by receiving end stations. As networks expand to support multiple local area networks, the true capability of multicast technology as an efficient means of transmission becomes more apparent.



As traffic is prepared to be forwarded over the physical network, it is necessary for hosts in shared collision domains to determine whether any traffic is currently occupying the transmission medium. Transmission media such as in the case of 10Base2 provides a shared medium over which CSMA/CD must be applied to ensure collisions are handled should they occur. If the transmission of a frame is detected on the link, the host will delay the forwarding of its own frames until such time as the line becomes available, following which the host will begin to forward frames from the physical interface towards the intended destination.

Where two hosts are connected over a medium capable of supporting full duplex transmission as in the case of media such as 10BaseT, it is considered not possible for transmitted frames to suffer collisions since transmission and receipt of frames occurs over separate wires and therefore there is no requirement for CSMA/CD to be implemented.

Frame Processing



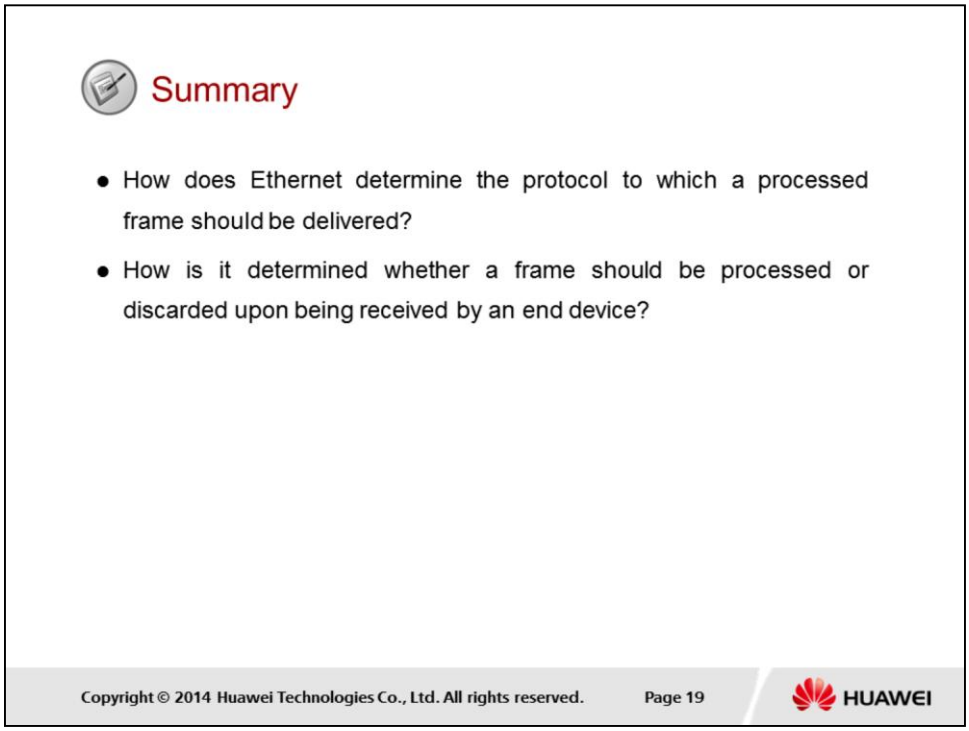
- Data link (frame) instructions are received, processed and discarded.

HUAWEI

Once a frame is forwarded from the physical interface of the host, it is carried over the medium to its intended destination. In the case of a shared network, the frame may be received by multiple hosts who will assess whether the frame is intended for their interface by analyzing the destination MAC address in the frame header. If the destination MAC address and the MAC address of the host are not the same, or the destination MAC address is not a MAC broadcast or multicast address to which the host is listening for, the frame will be ignored and discarded.

For the intended destination, the frame will be received and processed, initially by confirming that the frame is intended for the hosts physical interface. The host must also confirm that the integrity of the frame has been maintained during transmission by taking the value of the frame check sequence (FCS) field and comparing this value with a value determined by the receiving host. If the values do not match, the frame will be considered as corrupted and will be subsequently discarded.

For valid frames, the host will then need to determine the next stage of processing by analyzing the type field of the frame header and identify the protocol to which this frame is intended. In this example the frame type field contains a hexadecimal value of 0x0800 that identifies that the data taken from the frame should be forwarded to the Internet Protocol, prior to which, the frame header and trailer are discarded.



1. Data link layer frames contain a Type field that references the next protocol to which data contained within the frame should be forwarded. Common examples of forwarding protocols include IP (0x0800) and ARP (0x0806).
2. The destination MAC address contained within the frame header is analyzed by the receiving end station and compared to the MAC address associated with the interface on which the frame was received. If the destination MAC address and interface MAC address do not match, the frame is discarded.

Thank you

www.huawei.com

IP Addressing

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

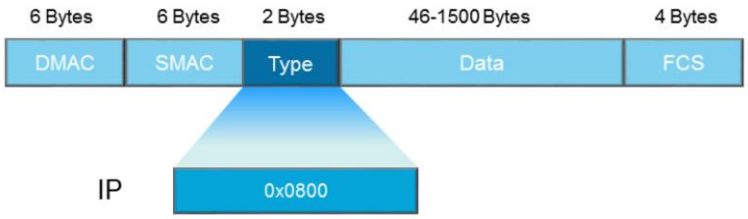
The Internet Protocol (IP) is designed to provide a means for internetwork communication that is not supported by lower layer protocols such as Ethernet. The implementation of logical (IP) addressing enables the Internet Protocol to be employed by other protocols for the forwarding of data in the form of packets between networks. A strong knowledge of IP addressing must be attained for effective network design along with clear familiarity of the protocol behavior, to support a clear understanding of the implementation of IP as a routed protocol.

Objectives

Upon completion of this section, trainees will be able to:

- Describe the fields and characteristics contained within IP.
- Distinguish between public, private and special IP address ranges.
- Successfully implement VLSM addressing.
- Explain the function of an IP gateway.

Next Header Processing

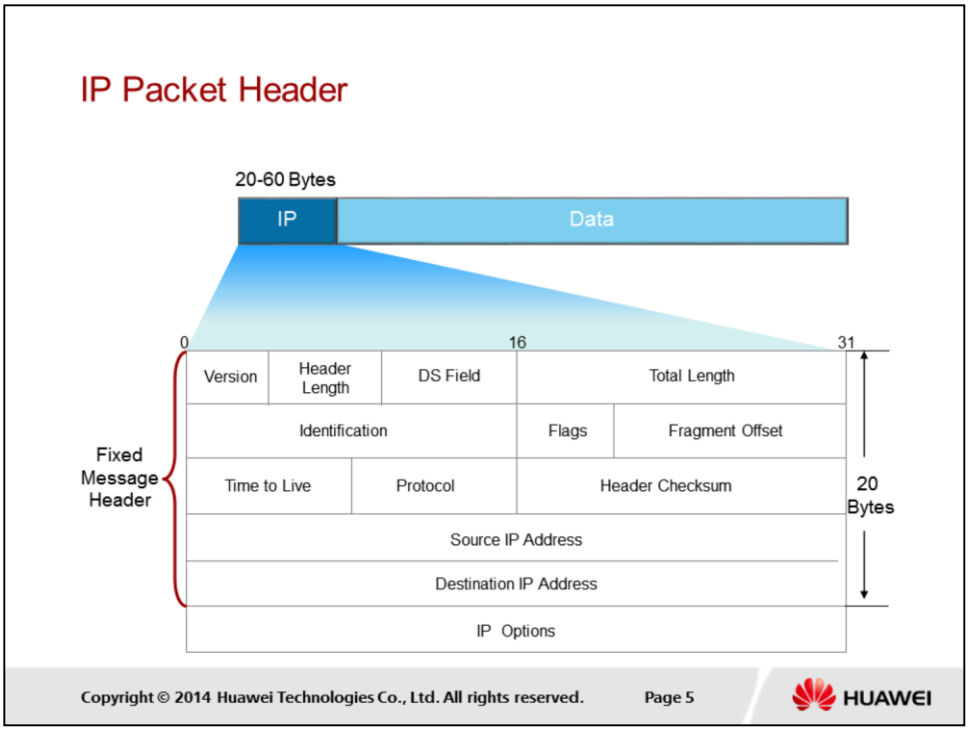


- The next set of instructions for processing are referenced in the type field of the frame header.

HUAWEI

Prior to discarding the frame header and trailer, it is necessary for the next set of instructions to be processed to be determined from the frame header. As highlighted, this is identified by determining the field value in the type field, which in the this instance represents a frame that is destined for the IP protocol following completion of the frame process.

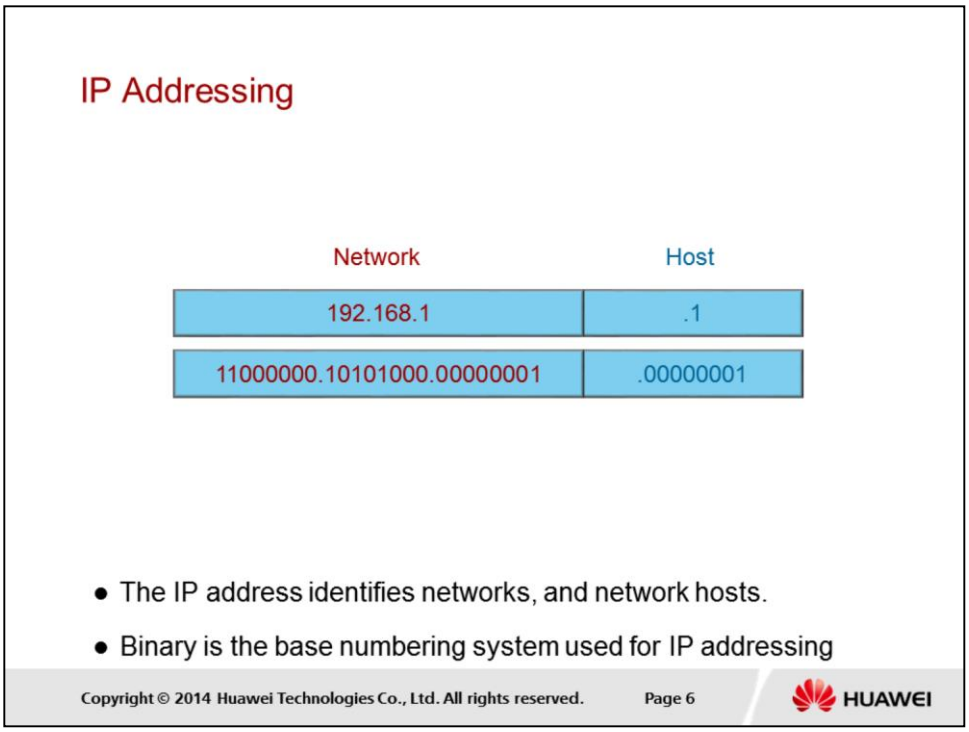
The key function of the frame is to determine whether the intended physical destination has been reached, that the integrity of the frame has remained in tact. The focus of this section will identify how data is processed following the discarding of the frame headers and propagation of the remaining data to the Internet Protocol.



The IP header is used to support two key operations, routing and fragmentation. Routing is the mechanism that allows traffic from a given network to be forwarded to other networks, since the data link layer represents a single network for which network boundaries exist. Fragmentation refers to the breaking down of data into manageable blocks that can be transmitted over the network.

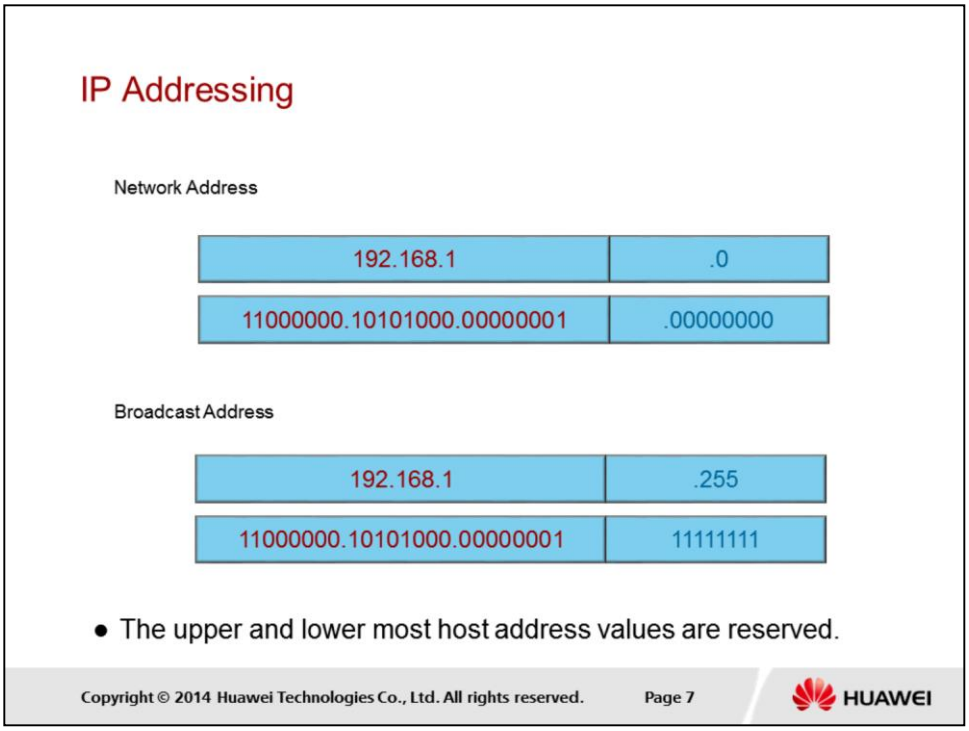
The IP header is carried as part of the data and represents an overhead of at least 20 bytes that references how traffic can be forwarded between networks, where the intended destination exists within a network different from the network on which the data was originally transmitted. The version field identifies the version of IP that is currently being supported, in this case the version is known as version four or IPv4. The DS field was originally referred to as the type of service field however now operates as a field for supporting differentiated services, primarily used as a mechanism for applying quality of service (QoS) for network traffic optimization, and is considered to be outside of the scope of this training.

The source and destination IP addressing are logical addresses assigned to hosts and used to reference the sender and the intended receiver at the network layer. IP addressing allows for assessment as to whether an intended destination exists within the same network or a different network as a means of aiding the routing process between networks in order to reach destinations beyond the local area network.



Each IPv4 address represents a 32 bit value that is often displayed in a dotted decimal format but for detailed understanding of the underlying behavior is also represented in a binary (Base 2) format. IP addresses act as identifiers for end systems as well as other devices within the network, as a means of allowing such devices to be reachable both locally and by sources that are located remotely, beyond the boundaries of the current network.

The IP address consists of two fields of information that are used to clearly specify the network to which an IP address belongs as well as a host identifier within the network range, that is for the most part unique within the given network.



Each network range contains two important addresses that are excluded from the assignable network range to hosts or other devices. The first of these excluded addresses is the network address that represents a given network as opposed to a specific host within the network. The network address is identifiable by referring to the host field of the network address, in which the binary values within this range are all set to 0, for which it should also be noted that an all 0 binary value may not always represent a 0 value in the dotted decimal notation.

The second excluded address is the broadcast address that is used by the network layer to refer to any transmission that is expected to be sent to all destinations within a given network. The broadcast address is represented within the host field of the IP address where the binary values within this range are all set to 1. Host addresses make up the range that exists between the network and broadcast addresses.

Decimal, Binary and Hexadecimal

Format	Value Range	Base Value
Binary	0 — 1	2
Decimal	0 — 9	10
Hexadecimal	0 — F	16

- Binary and Hexadecimal are common numbering systems used within IP networks.

The use of binary, decimal and hexadecimal notations are commonly applied throughout IP networks to represent addressing schemes, protocols and parameters, and therefore knowledge of the fundamental construction of these base forms is important to understanding the behavior and application of values within IP networks.

Each numbering system is represented by a different base value that highlights the number of values used as part of the base notations range. In the case of binary, only two values are ever used, 0 and 1, which in combination can provide for an increasing number of values, often represented as 2 to the power of x, where x denotes the number of binary values. Hexadecimal represents a base 16 notation with values ranging from 0 to F, (0-9 and A-F) where A represents the next value following 9 and F thus represents a value equivalent to 15 in decimal, or 1111 in binary.

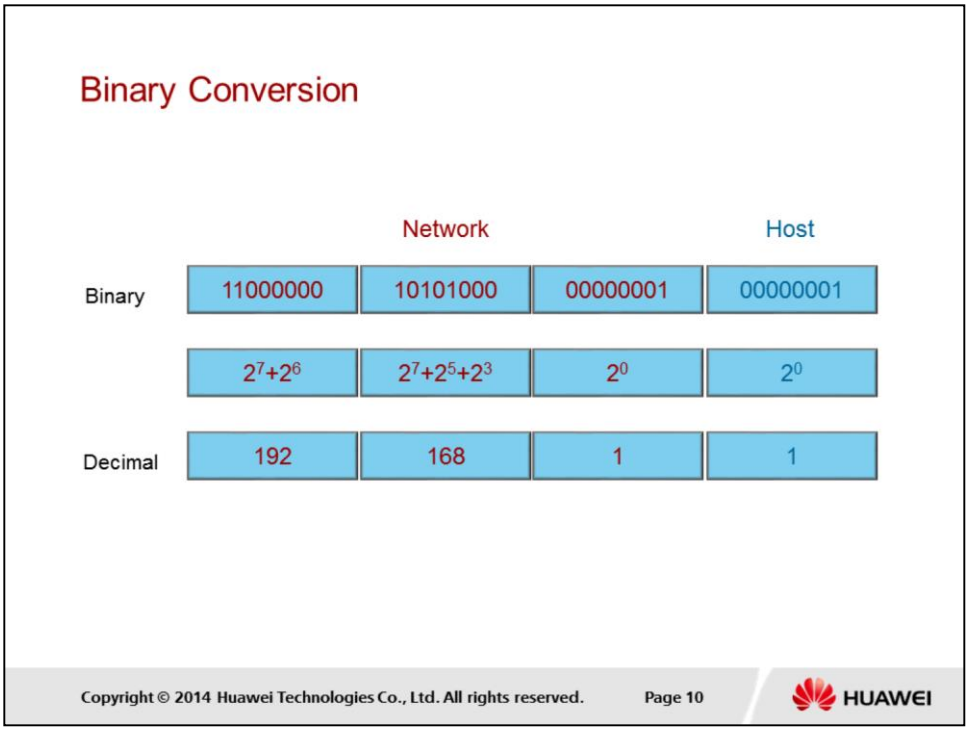
Binary vs. Decimal Conversion

Bit Order	1	1	1	1	1	1	1	1
Binary Power	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Binary	128	64	32	16	8	4	2	1

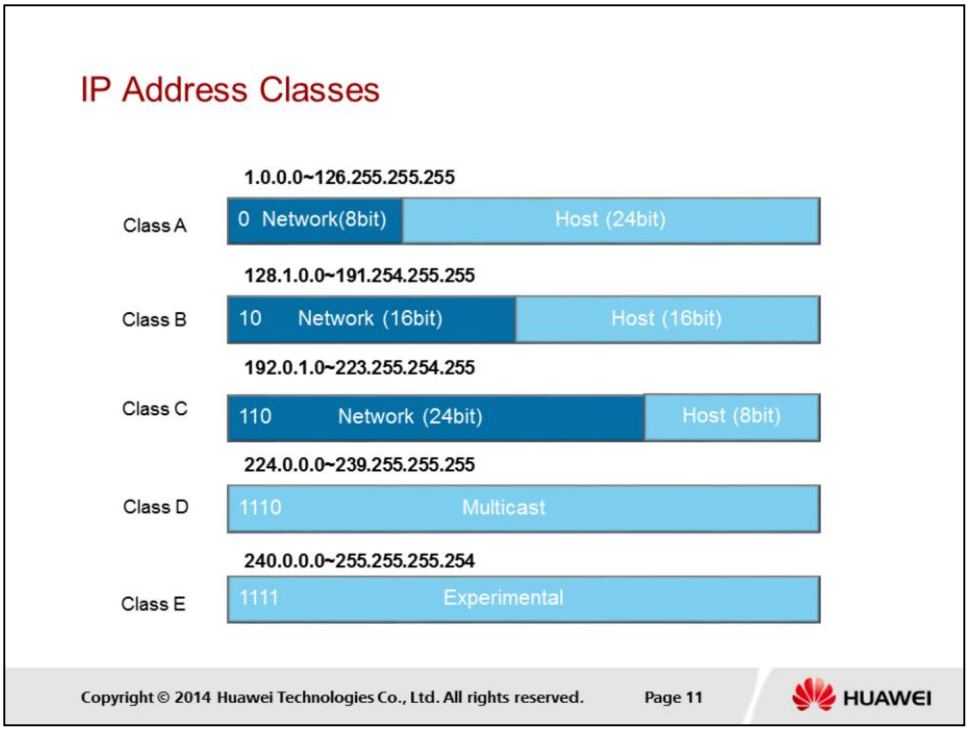
Decimal	Binary	Hexadecimal
0	00000000	00
1	00000001	01
2	00000010	02
3	00000011	03
4	00000100	04
5	00000101	05
6	00000110	06
7	00000111	07
8	00001000	08

Decimal	Binary	Hexadecimal
9	00001001	09
10	00001010	0A
11	00001011	0B
12	00001100	0C
13	00001101	0D
14	00001110	0E
15	00001111	0F
...	...	...
255	11111111	FF

A byte is understood to contain 8 bits and acts as a common notation within IP networks, thus a byte represents a bit value of 256, ranging from 0 through to 255. This information is clearly represented through translation of decimal notation to binary, and application of the base power to each binary value, to achieve the 256 bit value range. A translation of the numbering system for binary can be seen given in the example to allow familiarization with the numbering patterns associated with binary. The example also clearly demonstrates how broadcast address values in decimal, binary and hexadecimal are represented to allow for broadcasts to be achieved in both IP and MAC addressing at the network and data link layers.



The combination of 32 bits within an IP address correlates to four octets or bytes for which each can represent a value range of 256, giving a theoretical number of 4'294'967'296 possible IP addresses, however in truth only a fraction of the total number of addresses are able to be assigned to hosts. Each bit within a byte represents a base power and as such each octet can represent a specific network class, with each network class being based on either a single octet or a combination of octets. Three octets have been used as part of this example to represent the network with the fourth octet representing the host range that is supported by the network.



The number of octets supported by a network address is determined by address classes that break down the address scope of IPv4. Classes A, B and C are assignable address ranges, each of which supports a varied number of networks, and a number of hosts that are assignable to a given network. Class A for instance consist of 126 potential networks, each of which can support 2^{24}, or 16'777'216 potential host addresses, bearing in mind that network and broadcast addresses of a class range are not assignable to hosts.

In truth, a single Ethernet network could never support such a large number of hosts since Ethernet does not scale well, due in part to broadcasts that generate excessive network traffic within a single local area network. Class C address ranges allow for a much more balanced network that scales well to Ethernet networks, supplying just over 2 million potential networks, with each network capable of supporting around 256 addresses, of which 254 are assignable to hosts.

Class D is a range reserved for multicast, to allow hosts to listen for a specific address within this range, and should the destination address of a packet contain a multicast address for which the host is listening, the packet shall be processed in the same way as a packet destined for the hosts assigned IP address. Each class is easily distinguishable in binary by observing the bit value within the first octet, where a class A address for instance will always begin with a 0 for the high order bit, whereas in a Class B the first two high order bits are always set as 1 and 0, allowing all classes to be easily determined in binary.

IP Address Types

Private Address Ranges	
Class A	10.0.0.0~10.255.255.255
Class B	172.16.0.0~172.31.255.255
Class C	192.168.0.0~192.168.255.255

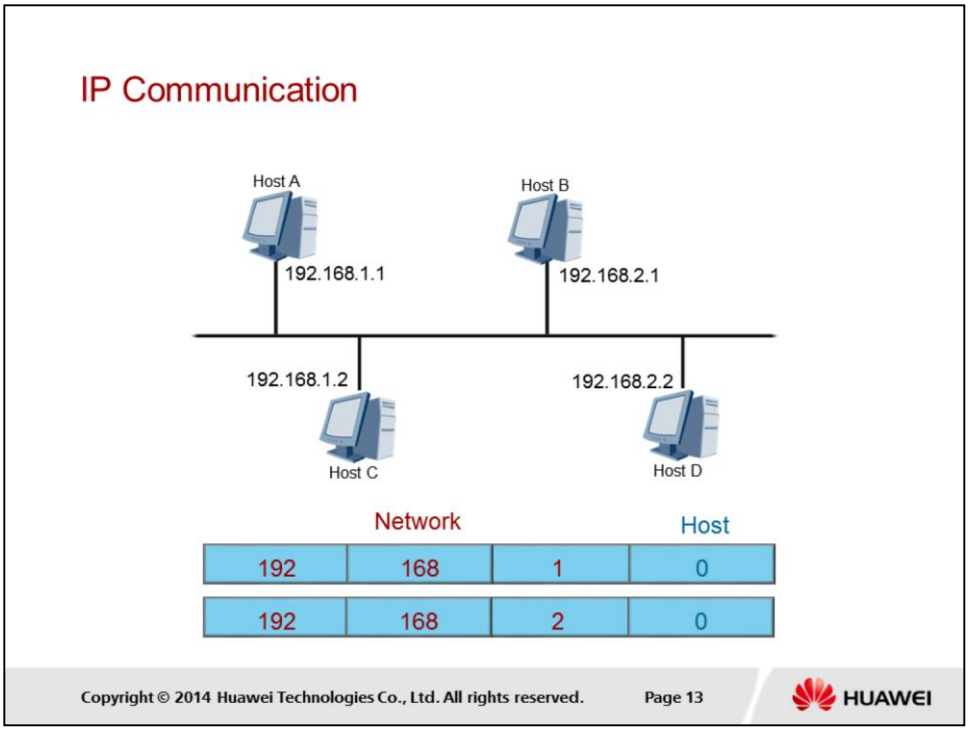
Special Addresses	
Diagnostic	127.0.0.0 ~ 127.255.255.255
Any Network	0.0.0.0
Network Broadcast	255.255.255.255

- The IP network address range has been divided, and certain addresses and ranges assigned special functions in the network.

Within IPv4, specific addresses and address ranges have been reserved for special purposes. Private address ranges exist within the class A, B and C address ranges to prolong the rapid decline in the number of available IP addresses. The number of actual end systems and devices that require IP addressing in the world today exceeds the 4’294’967’296 addresses of the 32 bit IPv4 address range, and therefore a solution to this escalating problem was to allocate private address ranges that could be assigned to private networks, to allow for conservation of public network addresses that facilitate communication over public network infrastructures such as the Internet.

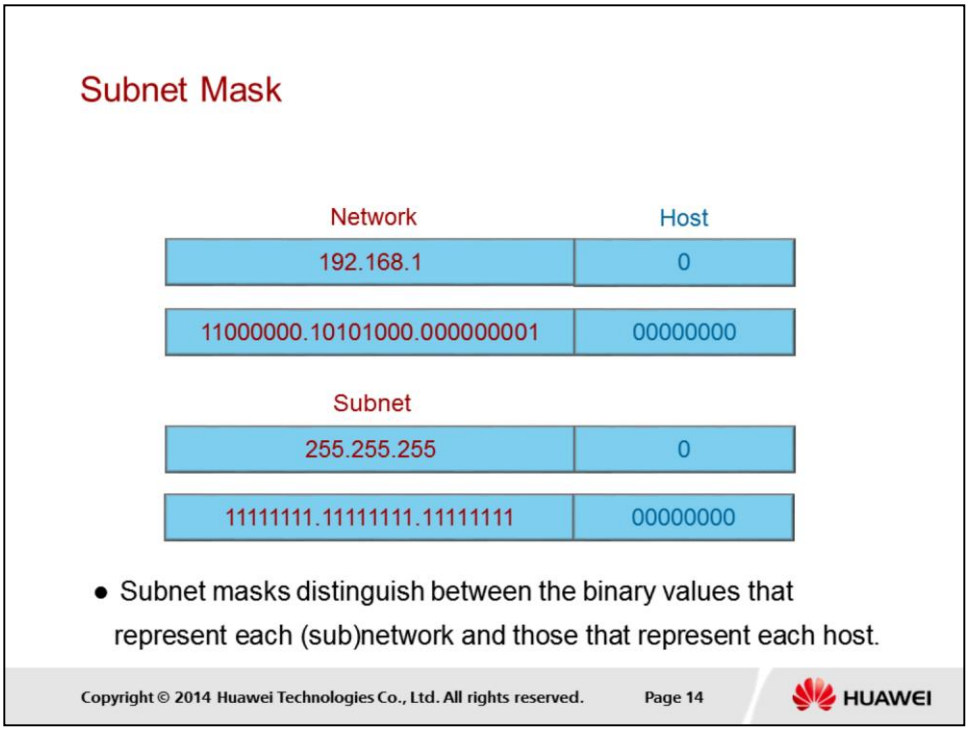
Private networks have become common throughout the enterprise network but hosts are unable to interact with the public network, meaning that address ranges can be reused in many disparate enterprise networks. Traffic bound for public networks however must undergo a translation of addresses before data can reach the intended destination.

Other special addresses include a diagnostic range denoted by the 127.0.0.0 network address, as well as the first and last addresses within the IPv4 address range, for which 0.0.0.0 represents any network and for which its application shall be introduced in more detail along with principles of routing. The address 255.255.255.255 represents a broadcast address for the IPv4 (0.0.0.0) network, however the scope of any broadcast in IP is restricted to the boundaries of the local area network from which the broadcast is generated.

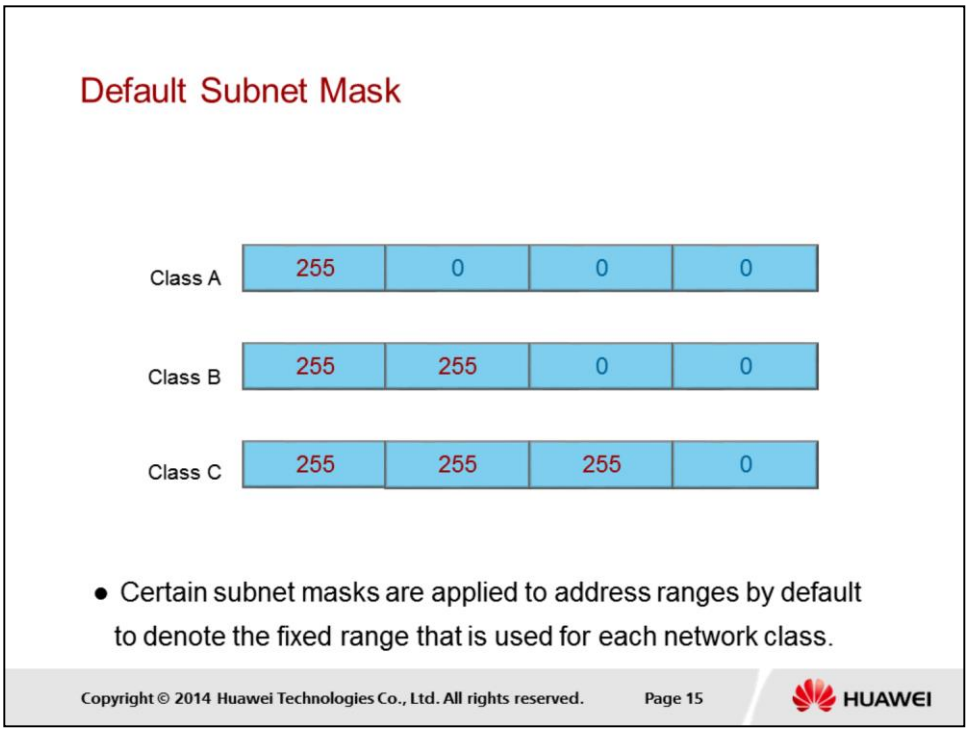


In order for a host to forward traffic to a destination, it is necessary for a host to have knowledge of the destination network. A host is naturally aware of the network to which it belongs but is not generally aware of other networks, even when those networks may be considered part of the same physical network. As such hosts will not forward data intended for a given destination until the host learns of the network and thus with it the interface via which the destination can be reached.

For a host to forward traffic to another host, it must firstly determine whether the destination is part of the same IP network. This is achieved through comparison of the destination network to the source network (host IP address) from which the data is originating. Where the network ranges match, the packet can be forwarded to the lower layers where Ethernet framing presides, for processing. In the case where the intended destination network varies from the originating network, the host is expected to have knowledge of the intended network and the interface via which a packet/frame should be forwarded before the packet can be processed by the lower layers. Without this information, the host will proceed to drop the packet before it even reaches the data link layer.

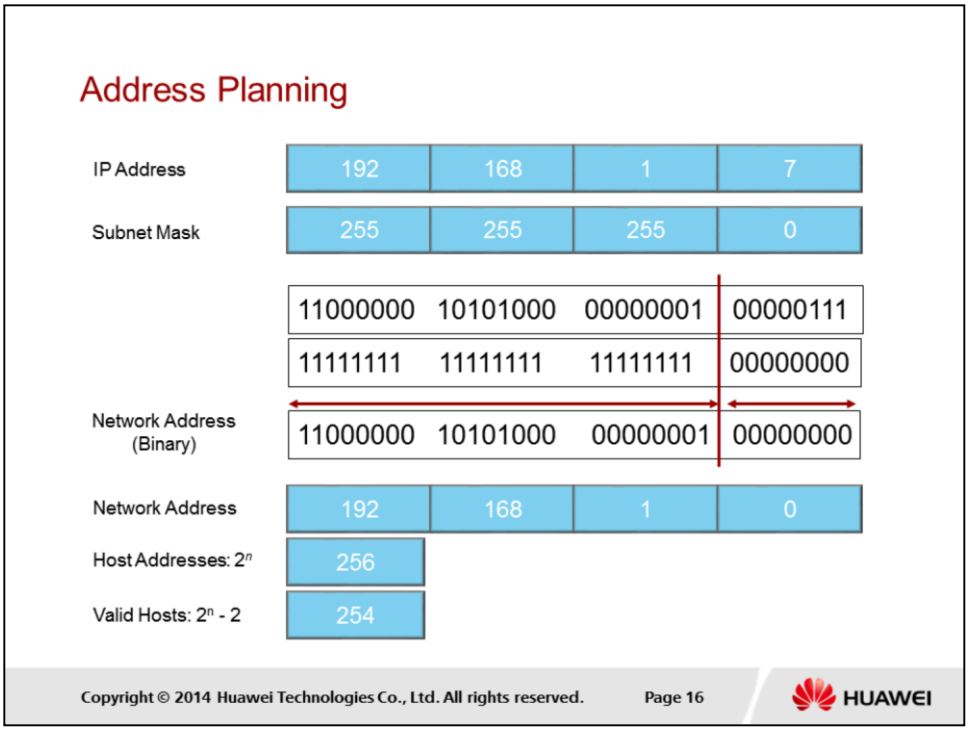


The identification of a unique network segment is governed by the implementation of a mask value that is used to distinguish the number of bits that represent the network segment, for which the remaining bits are understood as representing the number of hosts supported within a given network segment. A network administrator can divide a network address into sub-networks so that broadcast packets are transmitted within the boundaries of a single subnet. The subnet mask consists of a string of continuous and unbroken 1 values followed by an similar unbroken string of 0 values. The 1 values correspond to the network ID field whereas the 0 values correspond to the host ID field.



For each class of network address, a corresponding subnet mask is applied to specify the default size of the network segment. Any network considered to be part of the class A address range is fixed with a default subnet mask pertaining to 8 leftmost bits which comprise of the first octet of the IP address, with the remaining three octets remaining available for host ID assignment.

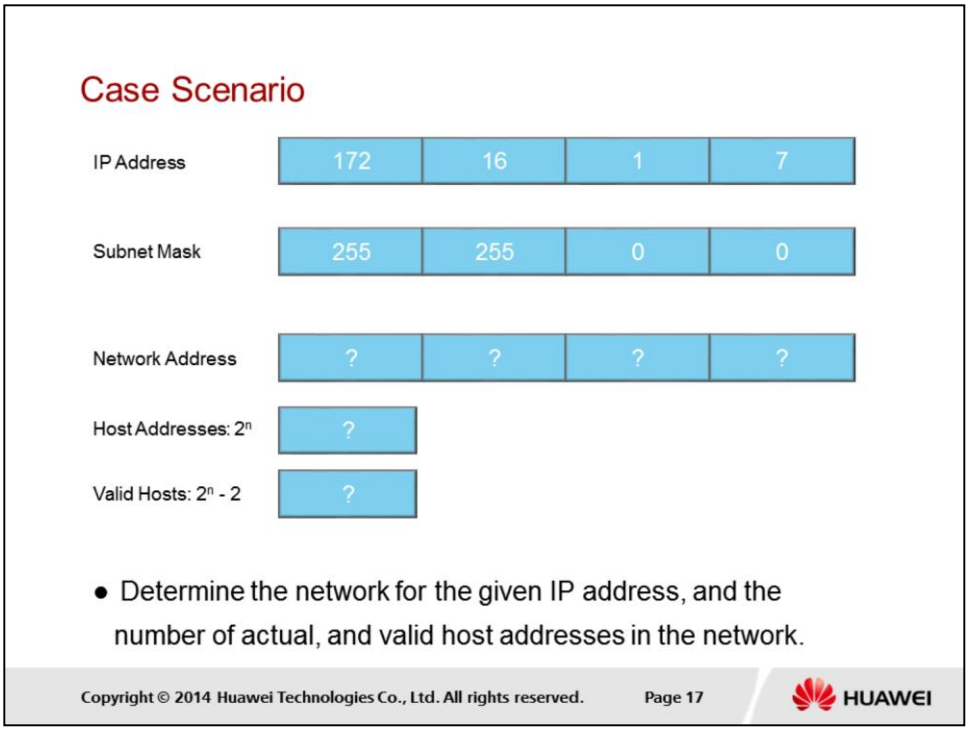
In a similar manner, the class B network reflects a default subnet mask of 16 bits, allowing a greater number of networks within the class B range at the cost of the number of hosts that can be assigned per default network. The class C network defaults to a 24 bit mask that provides a large number of potential networks but limits greatly the number of hosts that can be assigned within the default network. The default networks provide a common boundary to address ranges, however in the case of class A and class B address ranges, do not provide a practical scale for address allocation for Ethernet based networks.



Application of the subnet mask to a given IP address enables identification of the network to which the host belongs. The subnet mask will also identify the broadcast address for the network as well as the number of hosts that can be supported as part of the network range. Such information provides the basis for effective network address planning. In the example given, a host has been identified with the address of 192.168.1.7 as part of a network with a 24 bit default (class C) subnet mask applied. In distinguishing which part of the IP address constitutes the network and host segments, the default network address can be determined for the segment.

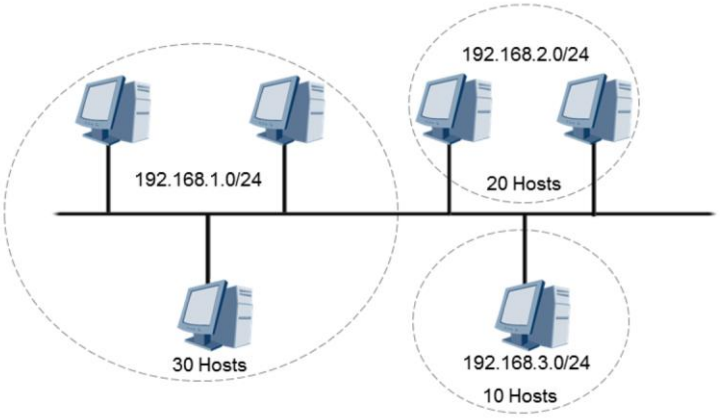
This is understood as the address where all host bit values are set to 0, in this case generating a default network address of 192.168.1.0. Where the host values are represented by a continuous string of 1 values, the broadcast address for the network can be determined. Where the last octet contains a string of 1 values, it represents a decimal value of 255, for which a broadcast address of 192.168.1.255 can be derived.

Possible host addresses are calculated based on a formula of 2^n where n represents the number of host bits defined by the subnet mask. In this instance n represents a value of 8 host bits, where 2^8 gives a resulting value of 256. The number of usable host addresses however requires that the network and broadcast addresses be deducted from this result to give a number of valid host addresses of 254.



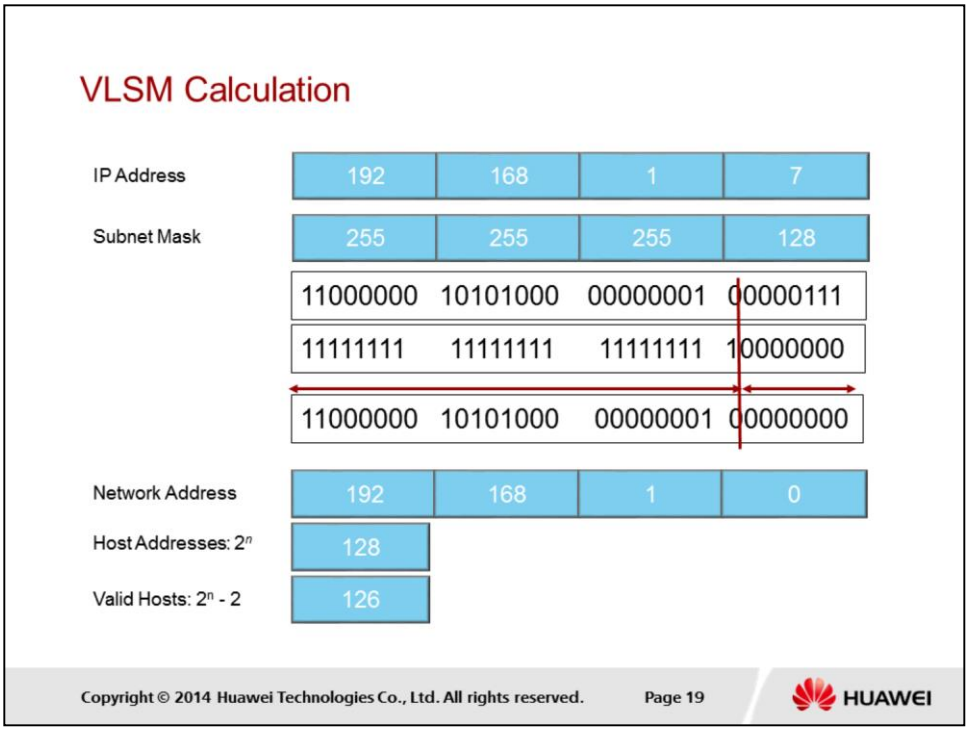
The case scenario provides a common class B address range to which it is necessary to determine the network to which the specified host belongs, along with the broadcast address and the number of valid hosts that are supported by the given network. Applying the same principles as with the class C address range, it is possible for the network address of the host to be determined, along with the range of hosts within the given network.

Addressing Limitations



- Network design using the default subnet mask results in address wastage

One of the main constraints of the default subnet mask occurs when multiple network address ranges are applied to a given enterprise in order to generate logical boundaries between the hosts within the physical enterprise network. The application of a basic addressing scheme may require a limited number of hosts to be associated with a given network, for which multiple networks are applied to provide the logical segmentation of the network. In doing so however, a great deal of address space remains unused, displaying the inefficiency of default subnet mask application.

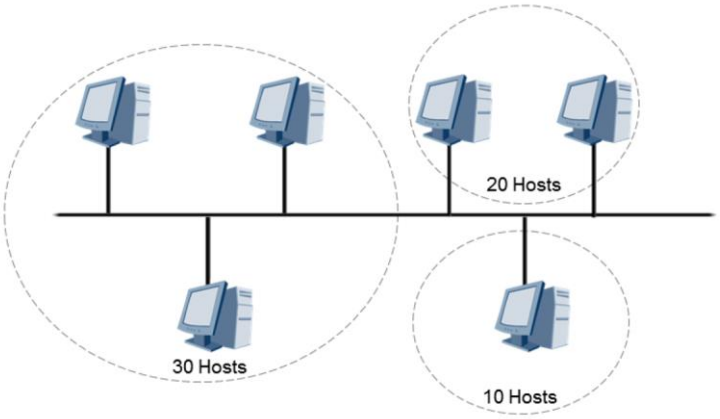


As a means of resolving the limitations of default subnet masks, the concept of variable length subnet masks are introduced, which enable a default subnet mask to be broken down into multiple sub-networks, which may be of a fixed length (a.k.a. fixed length subnet masks or FLSM) or of a variable length known commonly by the term VLSM. The implementation of such subnet masks consists of taking a default class based network and dividing the network through manipulation of the subnet mask.

In the example given, a simple variation has been made to the default class C network which by default is governed by a 24 bit mask. The variation comes in the form of a borrowed bit from the host ID which has been applied as part of the network address. Where the deviation of bits occurs in comparison to the default network, the additional bits represent what is known as the subnet ID.

In this case a single bit has been taken to represent the sub-network for which two sub-networks can be derived, since a single bit value can only represent two states of either 1 or 0. Where the bit is set to 0 it represents a value of 0, where is it is set to 1 it represents a value of 128. In setting the host bits to 0, the sub-network address can be found for each sub-network, by setting the host bits to 1, the broadcast address for each sub-network is identifiable. The number of supported hosts in this case represents a value of 2^7 minus the sub-network address and broadcast address for each sub-network, resulting in each sub-network supporting a total of 126 valid host addresses.

VLSM Case Scenario

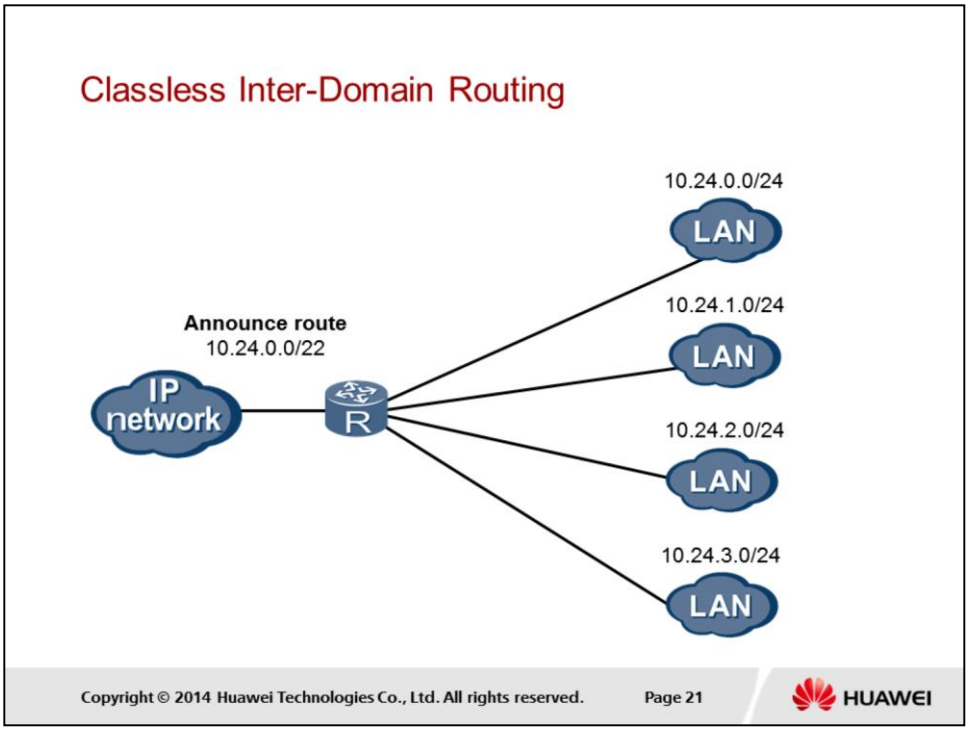


- Using only the network 192.168.1.0/24, implement VLSM for the given number of hosts in each network segment.

  HUAWEI

In relation to problem of address limitations in which default networks resulted in excessive address wastage, the concept of variable length subnet masks can be applied to reduce the address wastage and provide a more effective addressing scheme to the enterprise network.

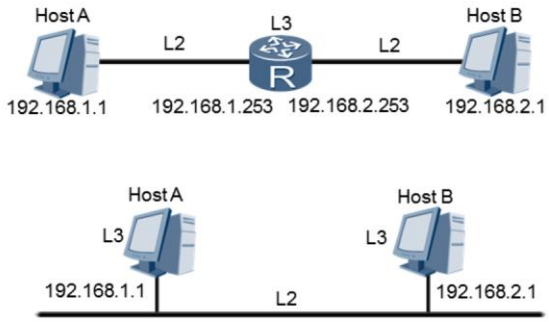
A single default class C address range has been defined, for which variable length subnet masks are required to accommodate each of the logical networks within a single default address range. Effective subnet mask assignment requires that the number of host bits necessary to accommodate the required number of hosts be determined, for which the remaining host bits can be applied as part of the subnet ID, that represents the variation in the network ID from the default network address.



Classless inter-domain routing was initially introduced as a solution to handle problems that were occurring as a result of the rapid growth of what is now known as the Internet. The primary concerns were to the imminent exhaustion of the class B address space that was commonly adopted by mid-sized organizations as the most suited address range, where class C was inadequate and where class A was too vast, and management of the 65534 host addresses could be achieved through VLSM. Additionally, the continued growth meant that gateway devices such as routers were beginning to struggle to keep up with the growing number of networks that such devices were expected to handle. The solution given involves transitioning to a classless addressing system in which classful boundaries were replaced with address prefixes.

This notation works on the principle that classful address ranges such as that of class C can be understood to have a 24 bit prefix that represents the subnet or major network boundary, and for which it is possible to summarize multiple network prefixes into a single larger network address prefix that represents the same networks but as a single address prefix. This has helped to alleviate the number of routes that are contained particularly within large scale routing devices that operate on a global scale, and has provided a more effective means of address management. The result of CIDR has had far reaching effects and is understood to have effectively slowed the overall exhaustion rate of the IPv4 address space.

IP Gateways



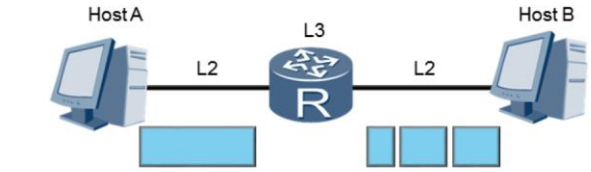
- Gateways use IP to forward packets between networks
- Hosts may act as gateways between networks in a LAN

The forwarding of packets requires that the packet first determine a forwarding path to a given network, and the interface via which a packet should be forwarded from, before being encapsulated as a frame and forwarded from the physical interface. In the case where the intended network is different from the originating network, the packet must be forwarded to a gateway via which the packet is able to reach it's intended destination.

In all networks, the gateway is a device that is capable of handling packets and making decisions as to how packets should be routed, in order to reach their intended destination. The device in question however must be aware of a route to the intended destination IP network before the routing of packets can take place. Where networks are divided by a physical gateway, the interface IP address (in the same network or sub-network) via which that gateway can be reached is considered to be the gateway address.

In the case of hosts that belong to different networks that are not divided by a physical gateway, it is the responsibility of the host to function as the gateway, for which the host must firstly be aware of the route for the network to which packets are to be forwarded, and should specify the hosts own interface IP address as the gateway IP address, via which the intended destination network can be reached.

IP Fragmentation



Version	Header Length	DS Field	Total Length
Identification		**Flags**	**Fragment Offset**
Time to Live	Protocol	Header Checksum	
Source IP Address			
Destination IP Address			
IP Options			

The data of forwarded packets exists in many formats and consists of varying sizes, often the size of data to be transmitted exceeds the size that is supported for transmission. Where this occurs it is necessary for the data block to be broken down into smaller blocks of data before transmission can occur. The process of breaking down this data into manageable blocks is known as fragmentation.

The identification, flags and fragment offset fields are used to manage reassembly of fragments of data once they are received at their final intended destination. Identification distinguishes between data blocks of traffic flows which may originate from the same host or different hosts. The flags field determines which of a number of fragments represents the last fragment at which time initiation of a timer is started prior to reassembly, and to notify that reassembly of the packet should commence.

Finally the fragment offset labels the bit value for each fragment as part of a number of fragments, the first fragment is set with a value of 0 and subsequent fragments specify the value of first bit following the previous fragment, for example where the initial fragment contains data bits 0 through to 1259, the following fragment will be assigned an offset value of 1260.

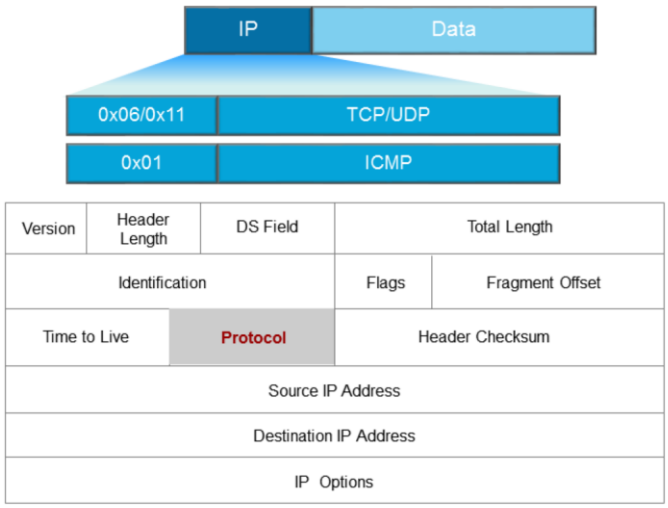
Time To Live

Version	Header Length	DS Field	Total Length
Identification		Flags	Fragment Offset
Time to Live	Protocol	Header Checksum	
Source IP Address			
Destination IP Address			
IP Options			

As packets are forwarded between networks, it is possible for packets to fall into loops where routes to IP networks have not been correctly defined within devices responsible for the routing of traffic between multiple networks. This can result in packets becoming lost within a cycle of packet forwarding that does not allow a packet to reach its intended destination. Where this occurs, congestion on the network will ensue as more and more packets intended for the same destination become subject to the same fate, until such time as the network becomes flooded with erroneous packets.

In order to prevent such congestion occurring in the event of such loops, a time to live (TTL) field is defined as part of the IP header, that decrements by a value of 1 each time a packet traverses a layer 3 device in order to reach a given network. The starting TTL value may vary depending on the originating source, however should the TTL value decrement to a value of 0, the packet will be discarded and an (ICMP) error message is returned to the source, based on the source IP address that can be found in the IP header of the wandering packet.

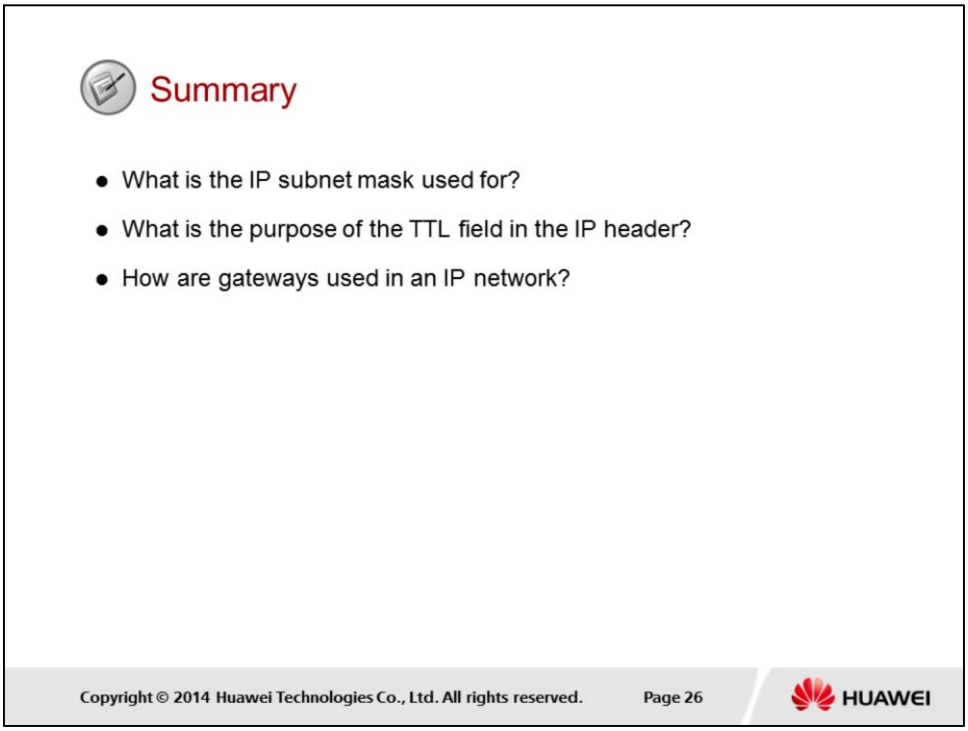
Protocol Field



HUAWEI

Upon verification that the packet has reached it's intended destination, the network layer must determine the next set of instructions that are to be processed. This is determined by analyzing the protocol field of the IP header. As with the type field of the frame header, a hexadecimal value is used to specify the next set of instructions to be processed.

It should be understood that the protocol field may refer to protocols at either the network layer, such as in the case of the Internet Control Message Protocol (ICMP), but may also refer to upper layer protocols such as the Transmission Control Protocol (06/0x06) or User Datagram Protocol (17/0x11), both of which exist as part of the transport layer within both the TCP/IP and OSI reference models.



1. The IP subnet mask is a 32 bit value that describes the logical division between the bit values of an IP address. The IP address is as such divided into two parts for which bit values represent either a network or sub-network, and the host within a given network or sub-network.
2. IP packets that are unable to reach the intended network are susceptible to being indefinitely forwarded between networks in an attempt to discover their ultimate destination. The Time To Live (TTL) feature is used to ensure that a lifetime is applied to all IP packets, so as to ensure that in the event that an IP packet is unable to reach it's destination, it will eventually be terminated. The TTL value may vary depending on the original source.
3. Gateways represent points of access between IP networks to which traffic can be redirected, or routed in the event that the intended destination network varies from the network on which the packet originated.

Thank you

www.huawei.com

Internet Control Message Protocol

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

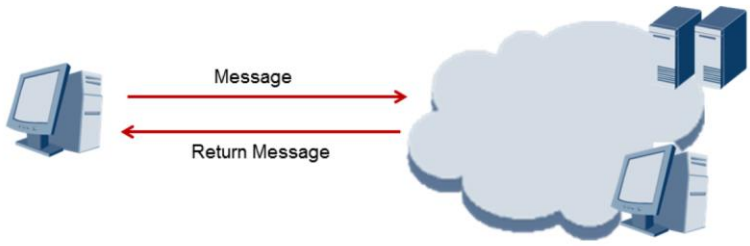
ICMP is a protocol that works alongside IP as a form of messaging protocol in order to compensate for the limited reliability of IP. The implementation of ICMP is required to be understood to familiarize with the behavior of numerous operations and applications that rely heavily on ICMP, in order to support underlying messaging, based on which further processes are often performed.

Objectives

Upon completion of this section, trainees will be able to:

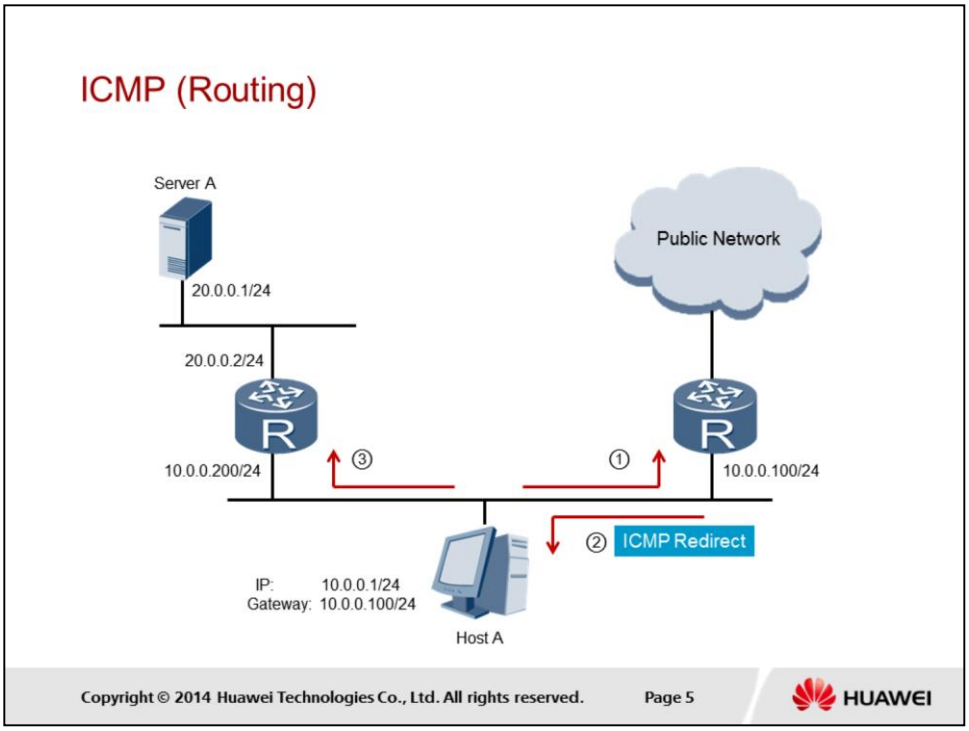
- Describe some of the processes to which ICMP is applied.
- Identify the common type and code values used in ICMP.
- Explain the function of ICMP in the ping and traceroute applications.

ICMP



- ICMP messages are used to support multiple operations including routing, diagnostics and errors.

The Internet Control Message Protocol is an integral part of IP designed to facilitate the transmission of notification messages between gateways and source hosts where requests for diagnostic information, support of routing, and as a means of reporting errors in datagram processing are needed. The purpose of these control messages is to provide feedback about problems in the communication environment, and does not guarantee that a datagram will be delivered, or that a control message will be returned.



ICMP Redirect messages represent a common scenario where ICMP is used as a means of facilitating routing functions. In the example, a packet is forwarded to the gateway by host A based on the gateway address of host A. The gateway identifies that the packet received is destined to be forwarded to the address of the next gateway which happens to be part of the same network as the host that originated the packet, highlighting a non optimal forwarding behavior between the host and the gateways.

In order to resolve this, a redirect message is sent to the host. The redirect message advises the host to send its traffic for the intended destination directly to the gateway to with which the destination network is associated, since this represents a shorter path to the destination. The gateway proceeds however to forward the data of the original packet to its intended destination.

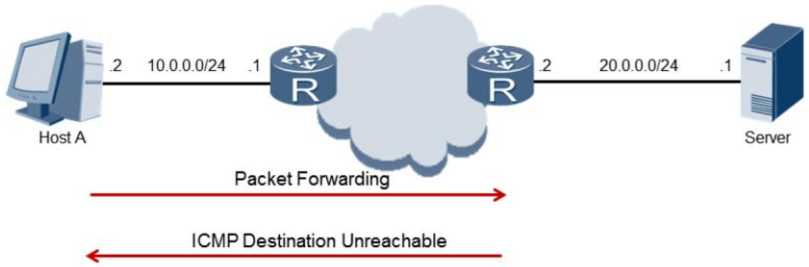
ICMP (Diagnostics)

- Two separate messages are used for the request and reply.
- Commonly associated with the Ping application.

HUAWEI

ICMP echo messages represent a means of diagnosis for determining primarily connectivity between a given source and destination, but also provides additional information such as the round trip time for transmission as a diagnostic for measuring delay. Data that is received in the echo message is returned as a separate echo reply message.

ICMP (Errors)



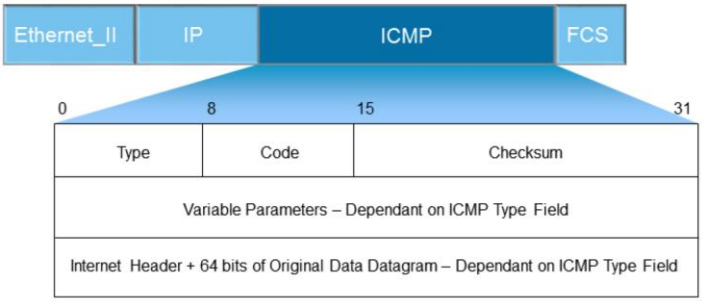
- Notifies the packet source of problems with packet forwarding.
- Uses the source IP address in the IP header for notification.

HUAWEI

ICMP provides various error reporting messages that often determine reachability issues and generate specific error reports that allow a clearer understanding from the perspective of the host as to why transmission to the intended destination failed.

Typical examples include cases where loops may have occurred in the network, and consequentially caused the time to live parameter in the IP header to expire, resulting in a “ttl exceeded in transit” error message being generated. Other examples include an intended destination being unreachable, which could relate to a more specific issue of the intended network not being known by the receiving gateway, or that the intended host within the destination network not being discovered. In all events an ICMP message is generated with a destination based on the source IP address found in the IP header, to ensure the message notifies the sending host.

ICMP Format



- ICMP parameters are represented in a type/code format.
- Additional data often carried to identify the undelivered packet.

HUAWEI

ICMP messages are sent using the basic IP header, which functions together as an integral part of the ICMP message, such is the case with the TTL parameter that is used to provide support for determining whether a destination is reachable. The format of the ICMP message relies on two fields for message identification in the form of a type/code format, where the type field provides a general description of the message type, and the code and a more specific parameter for the message type.

A checksum provides a means of validating the integrity of the ICMP message. An additional 32 bits are included to provide variable parameters, often unused and thus set as 0 when the ICMP message is sent, however in cases such as an ICMP redirect, the field contains the gateway IP address to which a host should redirect packets. The parameter field in the case of echo requests will contain an identifier and a sequence number, used to help the source associate sent echo requests with received echo replies, especially in the event multiple requests are forwarded to a given destination.

As a final means of tracing data to a specific process, the ICMP message may carry the IP header and a portion of the data that contains upper layer information that enables the source to identify the process for which an error occurred, such as cases where the ICMP TTL expires in transit.

ICMP Type & Code Fields

Type	Code	Description
0	0	Echo Reply
3	0	Network Unreachable
3	1	Host Unreachable
3	2	Protocol Unreachable
3	3	Port Unreachable
5	0	Redirect Datagram for the Network
8	0	Echo (Request)

- The *Type* value represents the format of a message.
- The *Code* value provides a more specific message description.

A wide number of ICMP type values exist to define clearly the different applications of the ICMP control protocol. In some cases the code field is not required to provide a more specific entry to the type field, as is found with echo requests that have a type field of 8 and the corresponding reply, which is generated and sent as a separate ICMP message to the source address of the sender, and defined using a type field of 0.

Alternatively, certain type fields define a very general type for which the variance is understood through the code field, as in the case of the type 3 parameter. A type field of 3 specifies that a given destination is unreachable, while the code field reflects the specific absence of either the network, host, protocol, port (TCP/UDP), ability to perform fragmentation (code 4), or source route (code 5) in which a packet, for which a forwarding path through the network is strictly or partially defined, fails to reach it’s destination.

ICMP Applications - Ping


```
<RTA>ping ?
-a       Select source IP address, the default is the IP address of
         the output interface
-c       Specify the number of echo requests to be sent, the
         default is 5
-n       Numeric output only. No attempt will be made to lookup
         host addresses for symbolic names
-t       Timeout in milliseconds to wait for each reply, the
         default is 2000ms
STRING<1-255> IP address or hostname of a remote system
.....
<RTA>ping 10.0.0.2
```


The application of ICMP can be understood through the use of tools such as Ping. The Ping application may be used as a tool in order to determine whether a destination is reachable as well as collect other related information. The parameters of the Ping application allow an end user to specify the behavior of the end system in generating ICMP messages, with consideration of the size of the ICMP datagram, the number of ICMP messages generated by the host, and also the duration in which it is expected a reply is received before a timeout occurs. This is important where a large delay occurs since a timeout may be reported by the Ping application before the ICMP message has had the opportunity to return to the source.

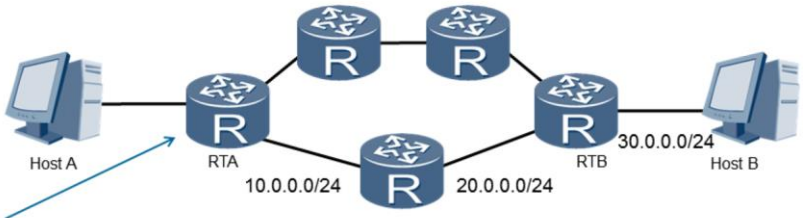
Ping Results

```
<RTA>ping 10.0.0.2
  PING 10.0.0.2 : 56  data bytes, press CTRL_C to break
    Reply from 10.0.0.2 : bytes=56 Sequence=1 ttl=255 time=340 ms
    Reply from 10.0.0.2 : bytes=56 Sequence=2 ttl=255 time=10 ms
    Reply from 10.0.0.2 : bytes=56 Sequence=3 ttl=255 time=30 ms
    Reply from 10.0.0.2 : bytes=56 Sequence=4 ttl=255 time=30 ms
    Reply from 10.0.0.2 : bytes=56 Sequence=5 ttl=255 time=30 ms

  --- 10.0.0.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 10/88/340 ms
```


The general output of an ICMP response to a Ping generated ICMP request details the destination to which the datagram was sent and the size of the datagram generated. In addition the sequence number of the sequence field that is carried as part of the echo reply (type 0) is displayed along with the TTL value that is taken from the IP header, as well as the round trip time which again is carried as part of the IP options field in the IP header.

ICMP Application – Traceroute




```
<RTA>tracert ?
-a       Set source IP address, the default is the IP
         address of the output interface
-f       First time to live, the default is 1
-m       Max time to live, the default is 30
-name    Display the host name of the router on each hop
-p       Destination UDP port number, the default is 33434
STRING<1-255> IP address or hostname of a remote system
.....................................................
<RTA>tracert 30.0.0.2
```


Another common application to ICMP is traceroute, which provides a means of measuring the forwarding path and delay on a hop by hop basis between multiple networks, through association with the TTL value within the IP header.

For a given destination, the reachability to each hop along the path is measured by initially defining a TTL value in the IP header of 1, causing the TTL value to expire before the receiving gateway is able to propagate the ICMP message any further, thus generating a TTL expired in transit message together with timestamp information, allowing for a hop by hop assessment of the path taken through the network by the datagram to the destination, and a measurement of the round trip time. This provides an effective means of identifying the point of any packet loss or delay that may be incurred in the network and also aids in the discovery of routing loops.

Traceroute Results

```
<RTA>tracert 30.0.0.2

traceroute to 30.0.0.2(30.0.0.2), max hops:30, packet length:40,
press CTRL_C to break

 1 10.0.0.2 130 ms  50 ms  40 ms

 2 20.0.0.2 80 ms  60 ms  80 ms

 3 30.0.0.2 80 ms  60 ms  70 ms
```

- Traceroute displays hop-by-hop transmission results.
- TTL value is used to define a hop limit for each set of results.

  HUAWEI

The implementation of traceroute in Huawei ARG3 series routers adopts the use of the UDP transport layer protocol to define a service port as the destination. Each hop sends three probe packets, for which the TTL value is initially set to a value of 1 and incremented after every three packets. In addition, a UDP destination port of 33434 is specified for the first packet and incremented for every successive probe packet sent. A hop by hop result is generated, allowing for the path to be determined, as well as for any general delay that may occur to be discovered.

This is achieved by measuring the duration between when the ICMP message was sent and when the corresponding TTL expired in transit ICMP error is received. When receiving a packet, the ultimate destination is unable to discover the port specified in the packet, and thus returns an ICMP Type 3, Code 3 (Port Unreachable) packet, and after three attempts the traceroute test ends. The test result of each probe is displayed by the source, in accordance with the path taken from the source to the destination. If a fault occurs when the trace route command is used, the following information may be displayed:

!H: The host is unreachable.

!N: The network is unreachable.

!: The port is unreachable.

!P: The protocol type is incorrect.

!F: The packet is incorrectly fragmented.

!S: The source route is incorrect.

Summary

- Which two ICMP message types are used as part of a successful Ping?
- In the event that the TTL value in the IP header of a datagram reaches zero, what action will be taken by the receiving gateway?

1. The Ping application uses the echo request message of type 8 to attempt to discover the destination. A separate echo reply message, defined by a type field of 0, is returned to the original source based on the source IP address in the IP header field.
2. In the event that the TTL value of an IP datagram reaches 0 before the datagram is able to reach the intended destination, the gateway device receiving the datagram will proceed to discard it and return an ICMP message to the source to notify that the datagram in question was unable to reach the intended destination. The specific reason will be defined by the code value to reflect for example whether the failure was due to a failure to discover the host, a port on the host or whether the service for a given protocol was not supported etc.

Thank you

www.huawei.com

Address Resolution Protocol

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

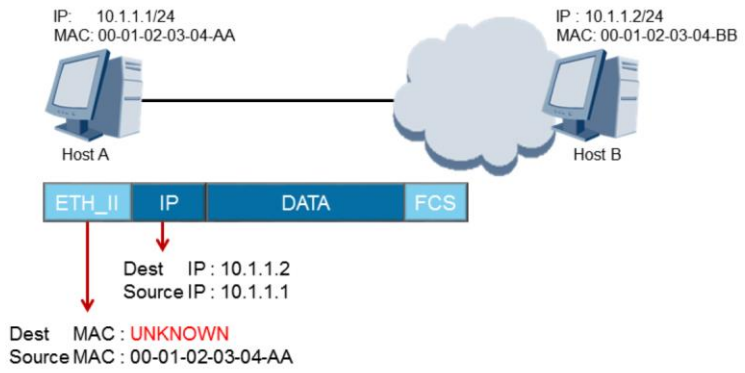
In order for data transmission to a network destination to be achieved it is necessary to build association between the network layer and lower layer protocols. The means by which the Address Resolution Protocol is used to build this association and prevent the unnecessary generation of additional broadcast traffic in the network should be clearly understood.

Objectives

Upon completion of this section, trainees will be able to:

- Explain how the MAC address is resolved using ARP.
- Explain the function of the ARP cache table.

ARP

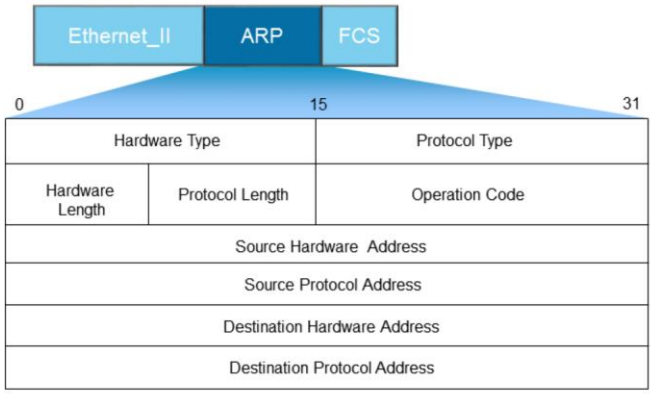


- Data link forwarding relies on knowledge of the MAC address of the data link layer destination.

HUAWEI

As data is encapsulated, the IP protocol at the network layer is able to specify the target IP address to which the data is ultimately destined, as well as the interface via which the data is to be transmitted, however before transmission can occur, the source must be aware of the target Ethernet (MAC) address to which data should be transmitted. The Address Resolution Protocol (ARP) represents a critical part of the TCP/IP protocol suite that enables discovery of MAC forwarding addresses to facilitate IP reachability. The Ethernet next hop must be discovered before data encapsulation can be completed.

ARP Format



- The ARP packet operates within the boundaries of the data link layer, as can be understood by the absence of an IP header.

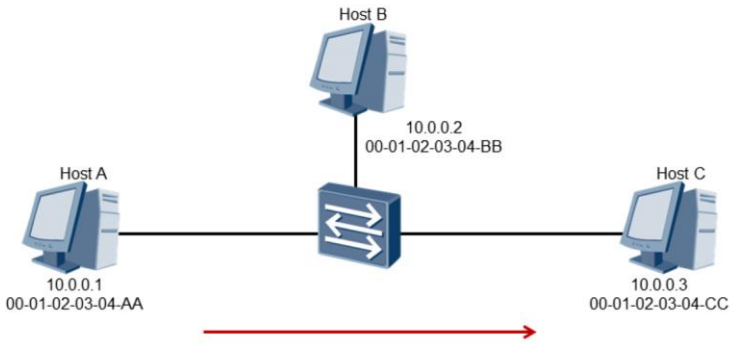
HUAWEI

The ARP packet is generated as part of the physical target address discovery process. Initial discovery will contain partial information since the destination hardware address or MAC address is to be discovered. The hardware type refers to Ethernet with the protocol type referring to IP, defining the technologies associated with the ARP discovery. The hardware and protocol length identifies the address length for both the Ethernet MAC address and the IP address, and is defined in bytes.

The operation code specifies one of two states, where the ARP discovery is set as REQUEST for which reception of the ARP transmission by the destination will identify that a response should be generated. The response will generate REPLY for which no further operation is necessary by the receiving host of this packet, and following which the ARP packet will be discarded. The source hardware address refers to the MAC address of the sender on the physical segment to which ARP is generated. The source protocol address refers to the IP address of the sender.

The destination hardware address specifies the physical (Ethernet) address to which data can be forwarded by the Ethernet protocol standards, however this information is not present in an ARP request, instead replaced by a value of 0. The destination protocol address identifies the intended IP destination for which reachability over Ethernet is to be established.

ARP Process

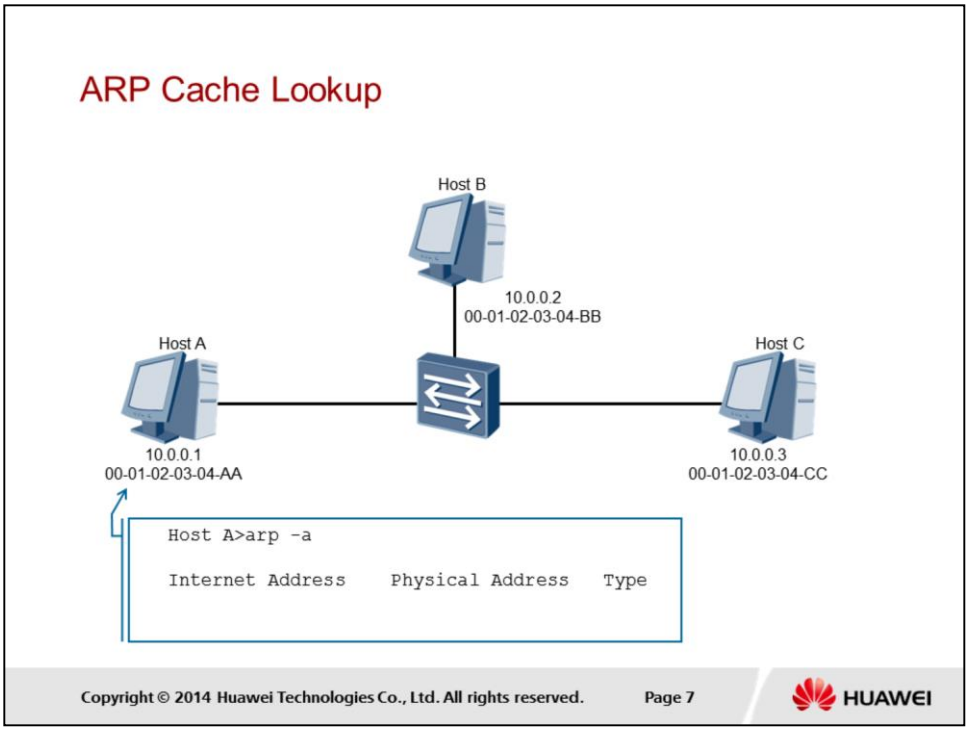


- Host A wishes to forward data to Host C, but must identify whether it is able to reach the destination at the data link layer.

HUAWEI

The network layer represents a logical path between a source and a destination. Reaching an intended IP destination relies on firstly being able to establish a physical path to the intended destination, and in order to do that, an association must be made between the intended IP destination and the physical next hop interface to which traffic can be forwarded.

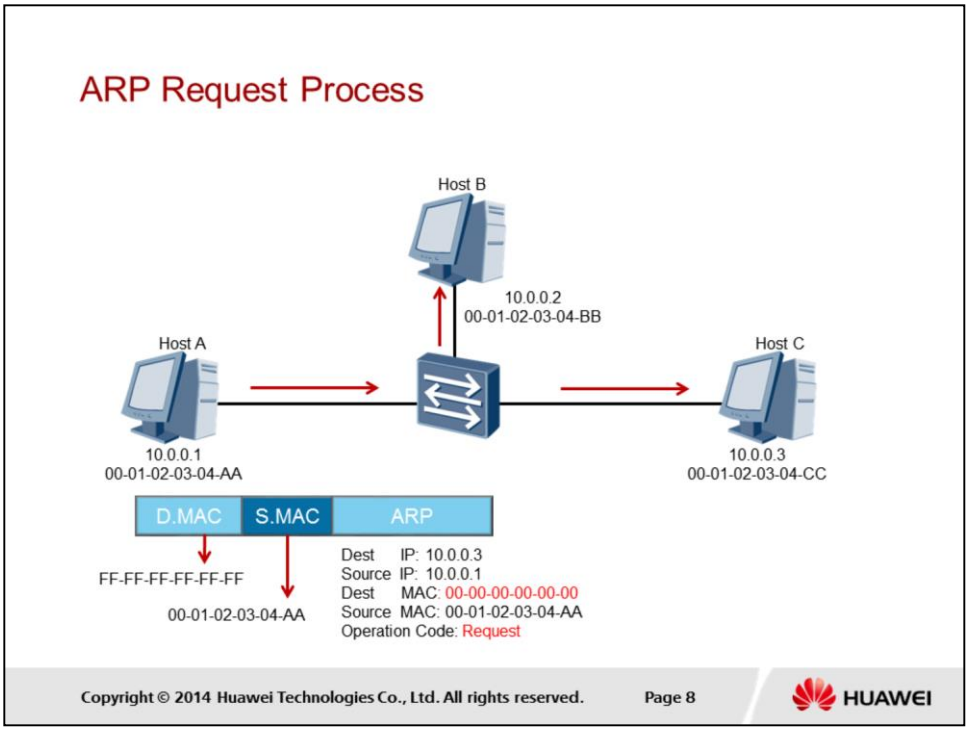
For a given destination the host will determine the IP address to which data is to be forwarded, however before encapsulation of the data can commence, the host must determine whether a physical forwarding path is known. If the forwarding path is known encapsulation to the destination can proceed, however quite often the destination is not known and ARP must be implemented before data encapsulation can be performed.



The ARP cache (pronounced as [kash]) is a table for association of host destination IP addresses and associated physical (MAC) addresses. Any host that is engaged in communication with a local or remote destination will first need to learn of the destination MAC via which communication can be established.

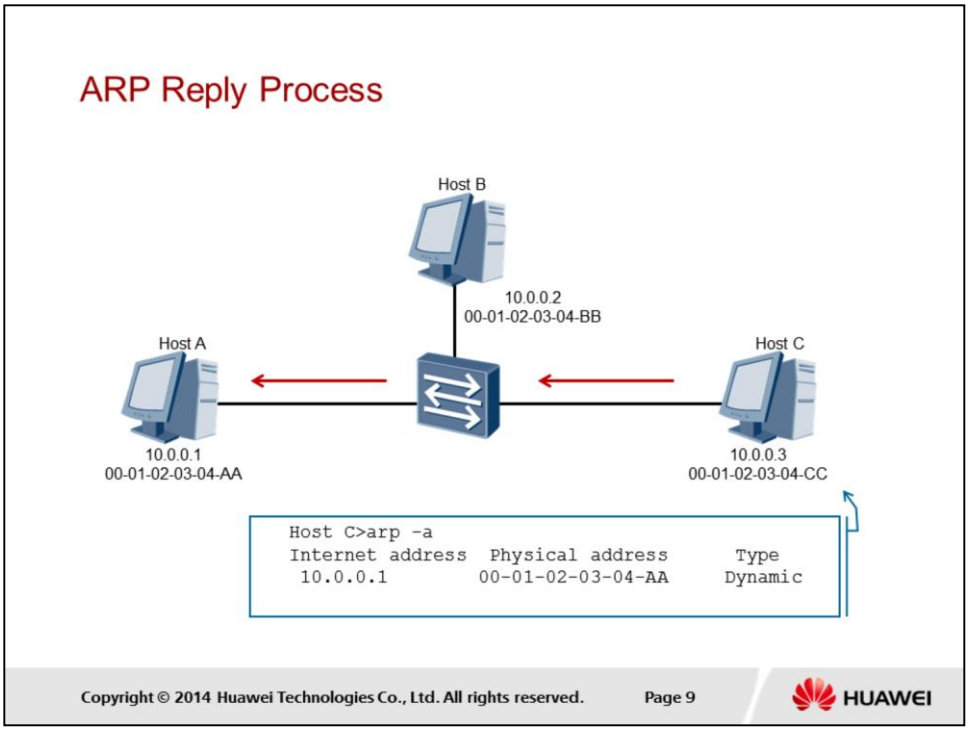
Learned addresses will populate the ARP cache table and remain active for a fixed period of time, during which the intended destination can be discovered without the need for addition ARP discovery processes. Following a fixed period, the ARP cache table will remove ARP entries to maintain the ARP cache table's integrity, since any change in the physical location of a destination host may result in the sending host inadvertently addressing data to a destination at which the destination host no longer resides.

The ARP cache lookup is the first operation that an end system will perform before determining whether it is necessary to generate an ARP request. For destinations beyond the boundaries of the hosts own network, an ARP cache lookup is performed to discover the physical destination address of the gateway, via which the intended destination network can be reached.



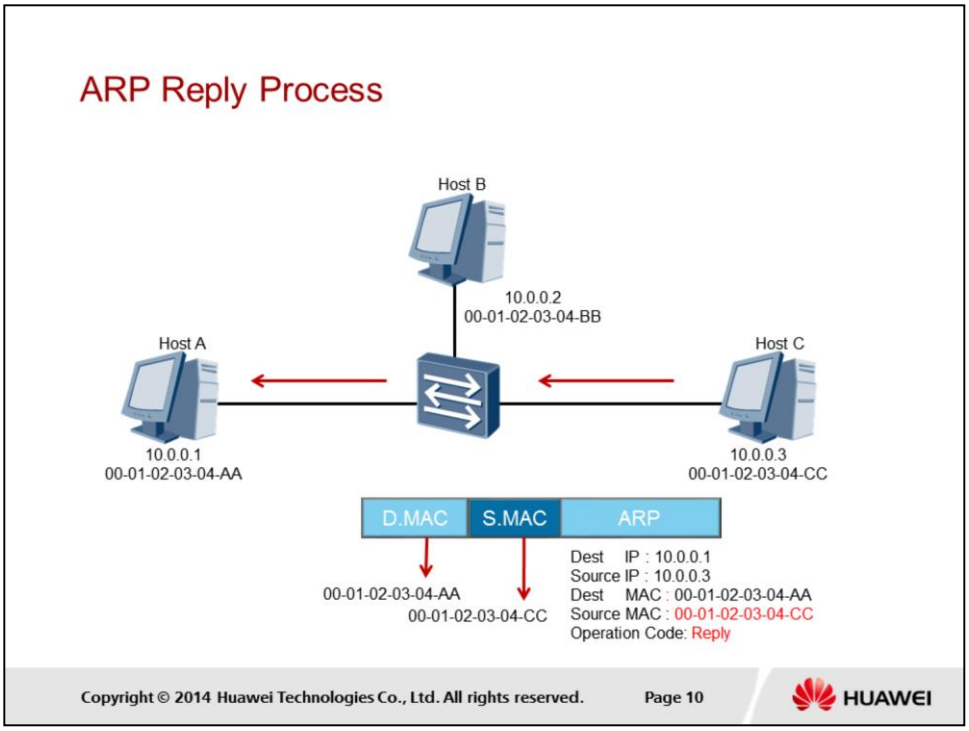
Where an ARP cache entry is unable to be determined, the ARP request process is performed. This process involves generation of an ARP request packet, and population of the fields with the source and destination protocol addresses, as well as the source hardware address. The destination hardware address is unknown. As such the destination hardware address is populated with a value equivalent to 0. The ARP request is encapsulated in an Ethernet frame header and trailer as part of the forwarding process. The source MAC address of the frame header is set as the source address of the sending host.

The host is currently unaware of the location of the destination and therefore must send the ARP request as a broadcast to all destinations within the same local network boundary. This means that a broadcast address is used as the destination MAC address. Once the frame is populated, it is forwarded to the physical layer where it is propagated along the physical medium to which the host is connected. The broadcasted ARP packet will be flooded throughout the network to all destinations including any gateway that may be present, however the gateway will prevent this broadcast from being forwarded to any network beyond the current network.



If the intended network destination exists, the frame will arrive at the physical interface of the destination at which point lower layer processing will ensue. ARP broadcasts mean that all destinations within the network boundary will receive the flooded frame, but will cease to process the ARP request, since the destination protocol address does not match to the IP address of those destinations.

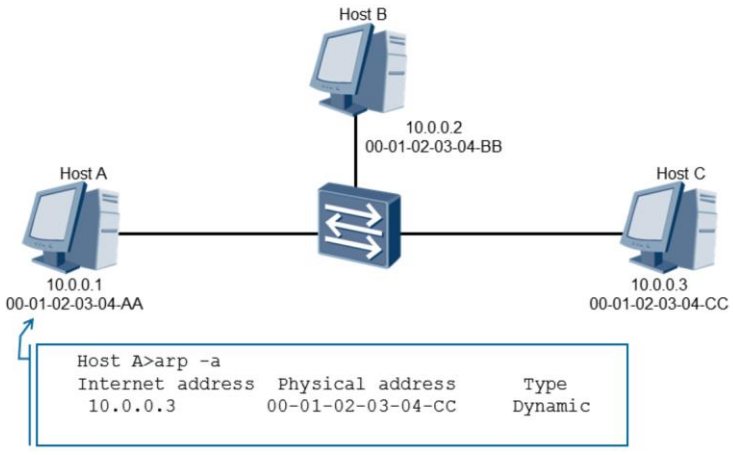
Where the destination IP address does match to the receiving host, the ARP packet will be processed. The receiving host will firstly process the frame header and then process the ARP request. The destination host will use the information from the source hardware address field in the ARP header to populate it's own ARP cache table, thus allowing for a unicast frame to be generated for any frame forwarding that may be required, to the source from which the ARP request was received.



The destination will determine that the ARP packet received is an ARP request and will proceed to generate an ARP reply that will be returned to the source, based on the information found in the ARP header. A separate ARP packet is generated for the reply, for which the source and destination protocol address fields will be populated. However, the destination protocol address in the ARP request packet now represents the source protocol address in the ARP reply packet, and similarly the source protocol address of the ARP request becomes the destination protocol address in the ARP reply.

The destination hardware address field is populated with the MAC of the source, discovered as a result of receiving the ARP request. For the required destination hardware address of the ARP request, it is included as the source hardware address of the ARP reply, and the operation code is set to reply, to inform the destination of the purpose of the received ARP packet, following which the destination is able to discard the ARP packet without any further communication. The ARP reply is encapsulated in the Ethernet frame header and trailer, with the destination MAC address of the Ethernet frame containing the MAC entry in the ARP cache table, allowing the frame to be forwarded as a unicast frame back to the host that originated the ARP request.

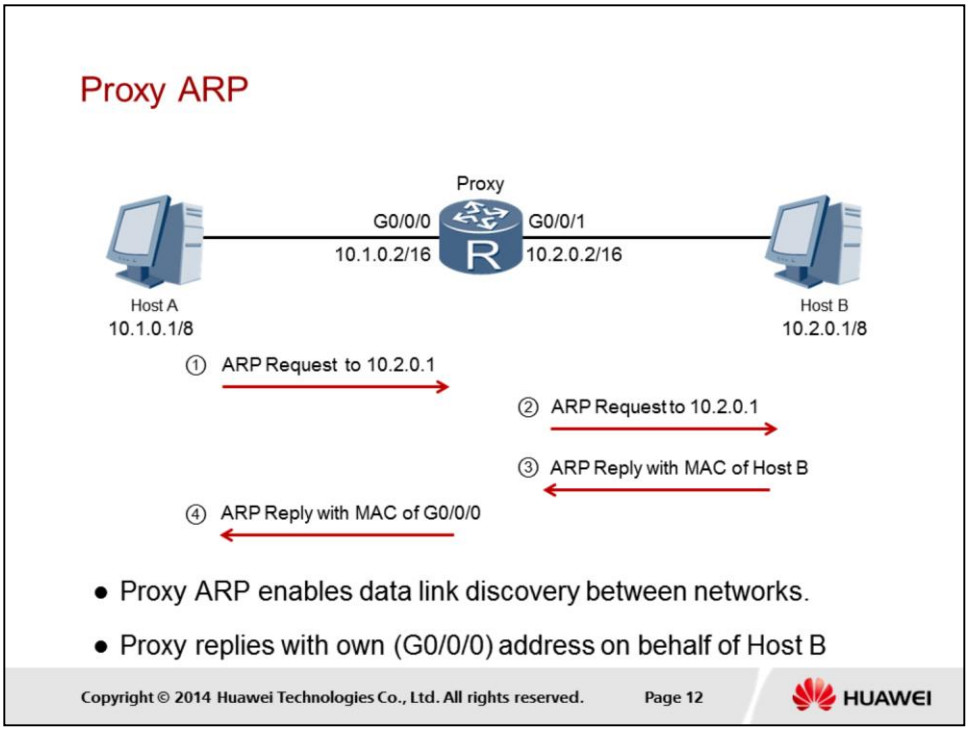
ARP Cache



HUAWEI

Upon receiving the ARP reply, the originating host will validate that the intended destination is correct based on the frame header, identify that the packet header is ARP from the type field and discard the frame headers. The ARP reply will then be processed, with the source hardware address of the ARP reply being used to populate the ARP cache table of the originating host (Host A).

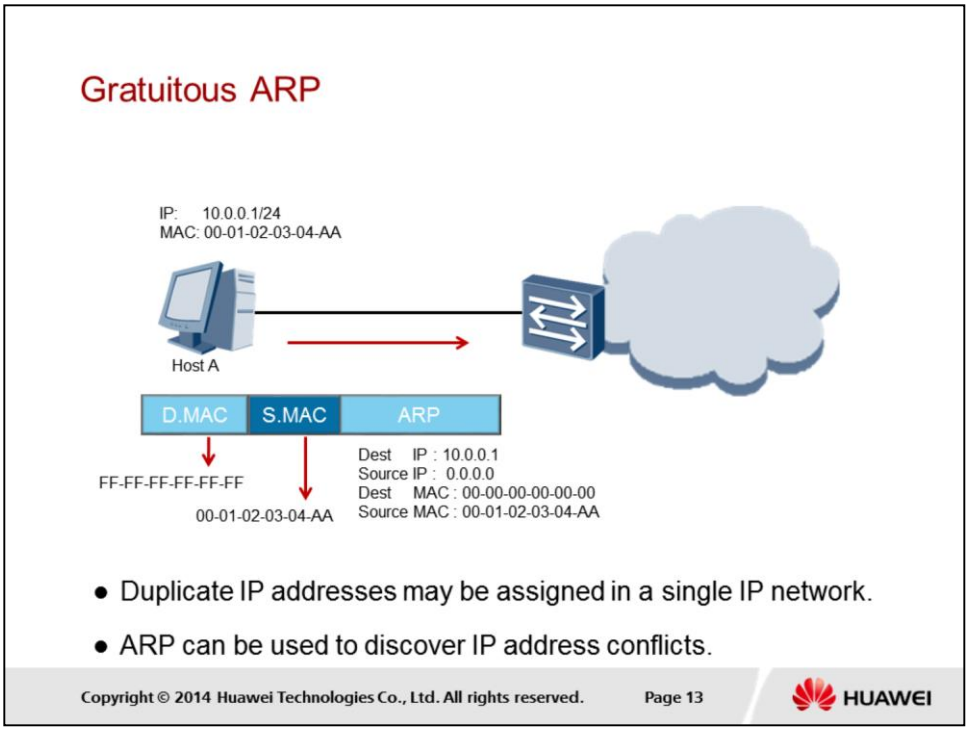
Following the processing of the ARP reply, the packet is discarded and the destination MAC information is used to facilitate the encapsulation process of the initial application or protocol that originally requested discovery of the destination at the data link layer.



The ARP protocol is also applied to other cases such as where transparent subnet gateways are to be implemented to facilitate communication across physical networks, where hosts are considered to be part of the same subnetwork. This is referred to as Proxy ARP since the gateway operates as a proxy for the two physical networks. When an ARP request is generated for a destination that is considered to be part of the same subnet, the request will eventually be received by the gateway. The gateway is able to determine that the intended destination exists beyond the physical network on which the ARP request was generated.

Since ARP requests cannot be forwarded beyond the boundaries of the broadcast domain, the gateway will proceed to generate its own ARP request to determine the reachability to the intended destination, using its own protocol and hardware addresses as the source addresses for the generated ARP request. If the intended destination exists, an ARP reply shall be received by the gateway for which the destinations source hardware address will be used to populate the ARP cache table of the gateway.

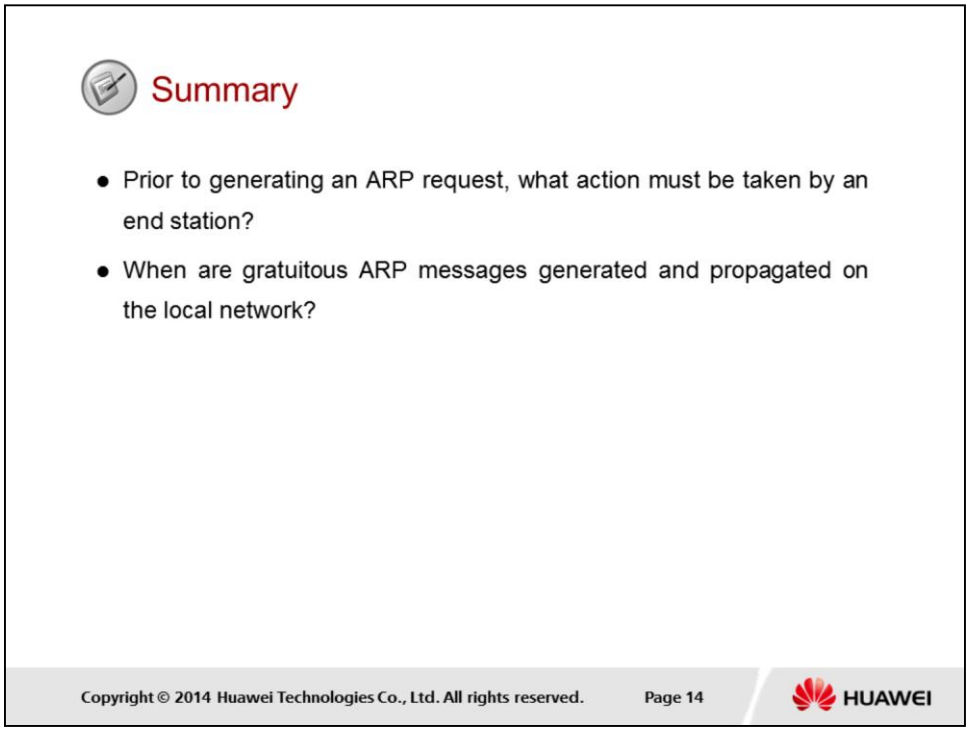
The gateway upon confirming the reachability to the intended destination will then generate an ARP reply to the original source (Host A) using the hardware address of the interface on which the ARP reply was forwarded. The gateway will as a result operate as an agent between the two physical networks to facilitate data link layer communication, with both hosts forwarding traffic intended for destinations in different physical networks to the relevant physical address of the “Proxy” gateway.



In the event that new hardware is introduced to a network, it is imperative that the host determine whether or not the protocol address to which it has been assigned is unique within the network, so as to prevent duplicate address conflicts. An ARP request is generated as a means of determining whether the protocol address is unique, by setting the destination address in the ARP request to be equal to the host's own IP address.

The ARP request is flooded throughout the network to all link layer destinations by setting the destination MAC as broadcast, to ensure all end stations and gateways receive the flooded frame. All destinations will process the frame, and should any destination discover that the destination IP address within the ARP request match the address of a receiving end station or gateway, an ARP reply will be generated and returned to the host that generated the ARP request.

Through this method the originating host is able to identify duplication of the IP address within the network, and flag an IP address conflict so to request that a unique address be assigned. This means of generating a request based on the hosts own IP address defines the basic principles of gratuitous ARP.



1. The host is required to initially determine whether it is already aware of a link layer forwarding address within its own ARP cache (MAC address table). If an entry is discovered the end system is capable of creating the frame for forwarding without the assistance of the address resolution protocol. If an entry cannot be found however, the ARP process will initiate, and an ARP request will be broadcasted on the local network.
2. Gratuitous ARP messages are commonly generated at the point in which an IP address is configured or changed for a device connected to the network, and at any time that a device is physically connected to the network. In both cases the gratuitous ARP process must ensure that the IP address that is used remains unique.

Thank you

www.huawei.com

Transport Layer Protocols

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

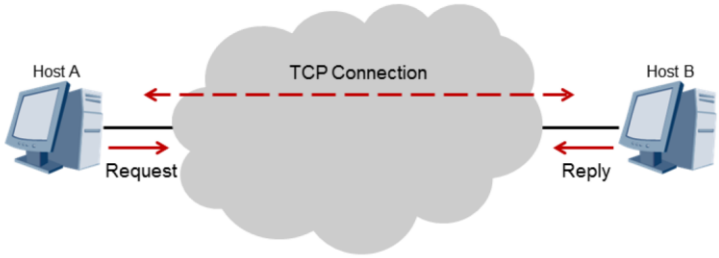
The transport layer is associated with the end-to-end behavior of transport layer protocols, that are defined once data reaches the intended destination. TCP and UDP represent the protocols commonly supported within IP networks. The characteristics of data, such as sensitivity to delay and the need for reliability often determines the protocols used at the transport layer. This section focuses on the knowledge of how such characteristics are supported through the behavior of each protocol.

Objectives

Upon completion of this section, trainees will be able to:

- Describe the common differences between TCP and UDP.
- Describe the forms of data to which TCP and UDP are applied.
- Identify well known TCP and UDP based port numbers.

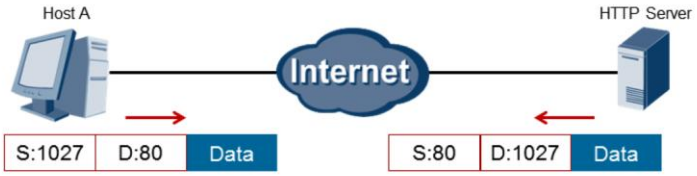
Transmission Control Protocol



- A connection is established before data is sent.

TCP is a connection-oriented, end-to-end protocol that exists as part of the transport layer of the TCP/IP protocol stack, in order to support applications that span over multi-network environments. The transmission control protocol provides a means of reliable inter-process communication between pairs of processes in host computers that are attached to distinct but interconnected computer communication networks. TCP relies on lower layer protocols to provide the reachability between process supporting hosts, over which a reliable connection service is established between pairs of processes. The connection-oriented behavior of TCP involves prior exchanges between the source and destination, through which a connection is established before transport layer segments are communicated.

TCP Ports



Protocol	Port
FTP	20 - 21
HTTP	80
TELNET	23
SMTP	25

- Ports represent individual services such as those listed above.

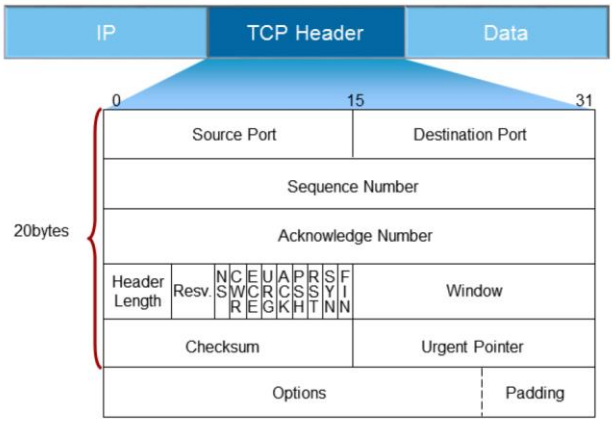
HUAWEI

As a means for allowing for many processes within a single host to use TCP communication facilities simultaneously, TCP provides a set of logical ports within each host. The port value together with the network layer address is referred to as a socket, for which a pair of sockets provide a unique identifier for each connection, in particular where a socket is used simultaneously in multiple connections. That is to say, a process may need to distinguish among several communication streams between itself and another process (or processes), for which each process may have a number of ports through which it communicates with the port or ports of other processes.

Certain processes may own ports and these processes may initiate connections on the ports that they own. These ports are understood as IANA assigned system ports or well known ports and exist in the port value range of 0 – 1023. A range of IANA assigned user or registered ports also exist in the range of 1024 – 49151, with dynamic ports, also known as private or ephemeral ports in the range of 49152 – 65535, which are not restricted to any specific application. Hosts will generally be assigned a user port value for which a socket is generated to a given application.

Common examples of TCP based applications for which well known port numbers have been assigned include FTP, HTTP, TELNET, and SMTP, which often will work alongside other well known mail protocols such as POP3 (port 110) and IMAP4 (port 143).

TCP Header



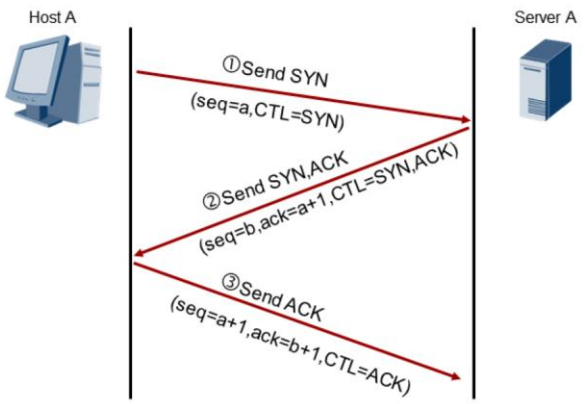
HUAWEI

The TCP header allows TCP based applications to establish connection-oriented data streams that are delivered reliability, and to which flow control is applied. A source port number is generated where a host intends to establish a connection with a TCP based application, for which the destination port will relate to a well known/registered port to which a well known/registered application is associated.

Code bits represent functions in TCP, and include an urgent bit (URG) used together the urgent pointer field for user directed urgent data notifications, acknowledgment of received octets in association with the acknowledgement field (ACK), the push function for data forwarding (PSH), connection reset operations (RST), synchronization of sequence numbers (SYN) and indication that no more data is to be received from the sender (FIN). Additional code bits were introduced in the form of ECN-Echo (ECE) and Congestion Window Reduced (CWR) flags, as a means of supporting congestion notification for delay sensitive TCP applications.

The explicit congestion notification (ECN) nonce sum (NS) was introduced as a follow-up alteration to eliminate the potential abuse of ECN where devices along the transmission path may remove ECN congestion marks. The Options field contains parameters that may be included as part of the TCP header, often used during the initial connection establishment, as in the case of the maximum segment size (MSS) value, that may be used to define the size of the segment that the receiver should use. TCP header size must be a sum of 32 bits, and where this is not the case, padding of 0 values will be performed.

TCP Connection Establishment



- A TCP connection is established after a three-way handshake.

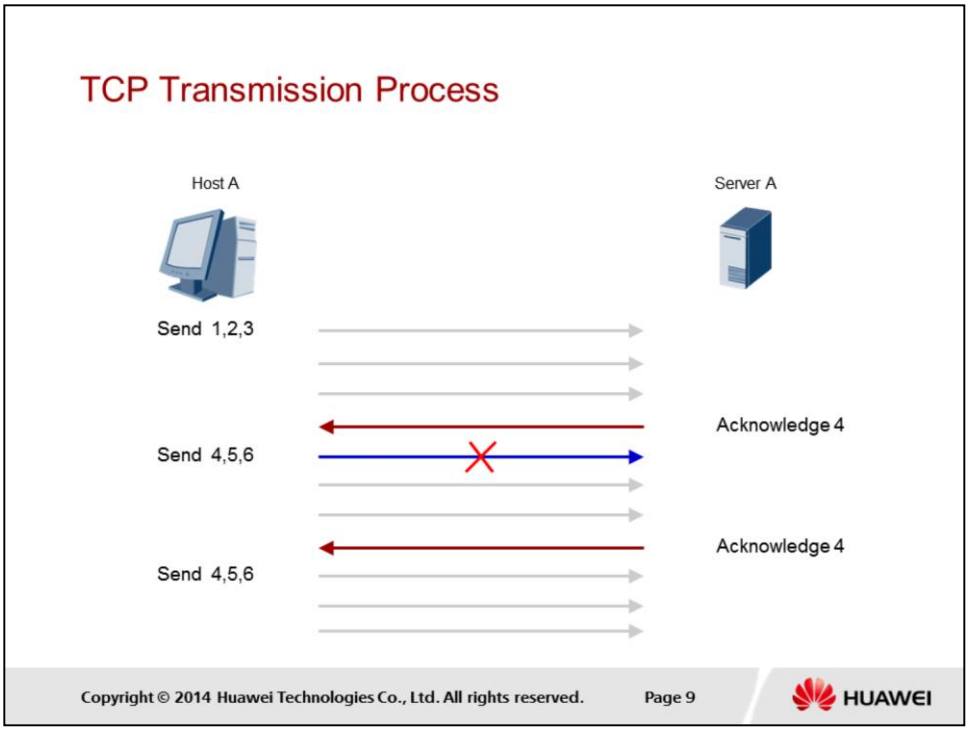
HUAWEI

When two processes wish to communicate, each TCP must first establish a connection (initialize the synchronization of communication on each side). When communication is complete, the connection is terminated or closed to free the resources for other uses. Since connections must be established between unreliable hosts and over the unreliable Internet domain, a handshake mechanism with clock-based sequence numbers is used to avoid erroneous initialization of connections.

A connection progresses through a series of states during establishment. The LISTEN state represents a TCP waiting for a connection request from any remote TCP and port. SYN-SENT occurs after sending a connection request and before a matching request is received. The SYN-RECEIVED state occurs while waiting for a confirming connection request acknowledgment, after having both received and sent a connection request. The ESTABLISHED state occurs following the handshake at which time an open connection is created, and data received can be delivered to the user.

The TCP three-way handshake mechanism begins with an initial sequence number being generated by the initiating TCP as part of the synchronization (SYN) process. The initial TCP segment is then set with the SYN code bit, and transmitted to the intended IP destination TCP to achieve a SYN-SENT state. As part of the acknowledgement process, the peering TCP will generate an initial sequence number of its own to synchronize the TCP flow in the other direction. This peering TCP will transmit this sequence number, as well as an acknowledgement number that equals the received sequence number incremented by one, together with set SYN and ACK code bits in the TCP header to achieve a SYN-RECEIVED state.

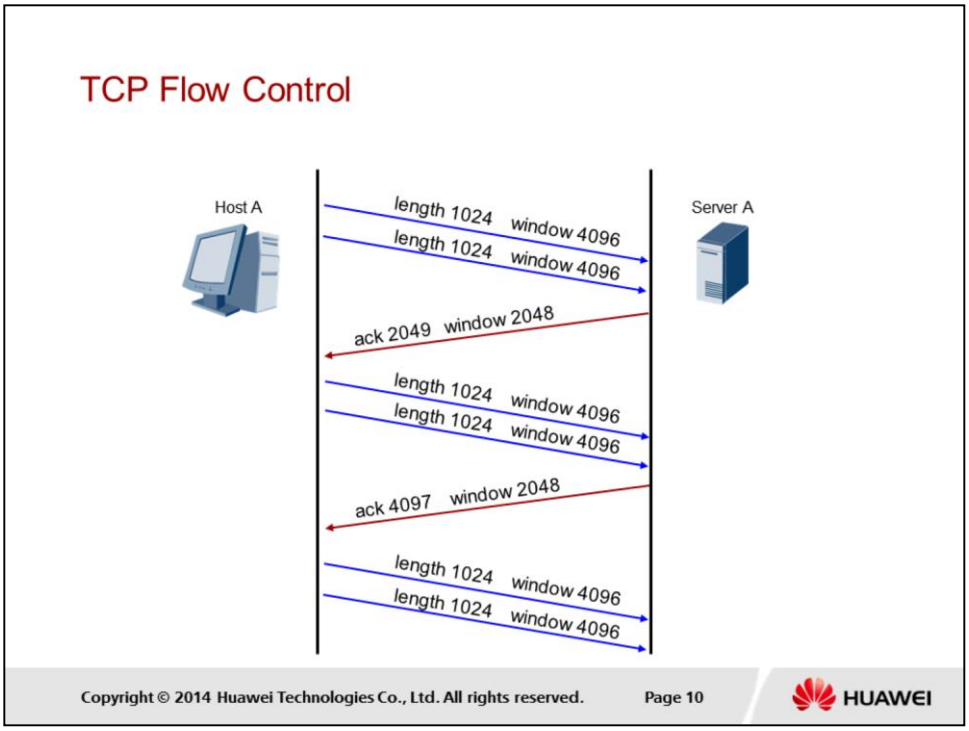
The final step of the connection handshake involves the initial TCP acknowledging the sequence number of the peering TCP by setting the acknowledgement number to equal the received sequence number plus one, together with the ACK bit in the TCP header, allowing an ESTABLISHED state to be reached.



Since the TCP transmission is sent as a data stream, every octet can be sequenced, and therefore each octet can be acknowledged. The acknowledgement number is used to achieve this by responding to the sender as confirmation of receipt of data, thus providing data transport reliability. The acknowledgement process however is cumulative, meaning that a string of octets can be acknowledged by a single acknowledgement by reporting to the source the sequence number that immediately follows the sequence number that was successfully received.

In the example a number of bytes (octets) are transmitted together before TCP acknowledgement is given. Should an octet fail to be transmitted to the destination, the sequence of octets transmitted will only be acknowledged to the point at which the loss occurred. The resulting acknowledgement will reflect the octet that was not received in order to reinitiate transmission from the point in the data stream at which the octet was lost.

The ability to cumulate multiple octets together before an acknowledgement enables TCP to operate much more efficiently, however a balance is necessary to ensure that the number of octets sent before an acknowledgement is required is not too extreme, for if an octet fails to be received, the entire stream of octets from the point of the loss must be retransmitted.

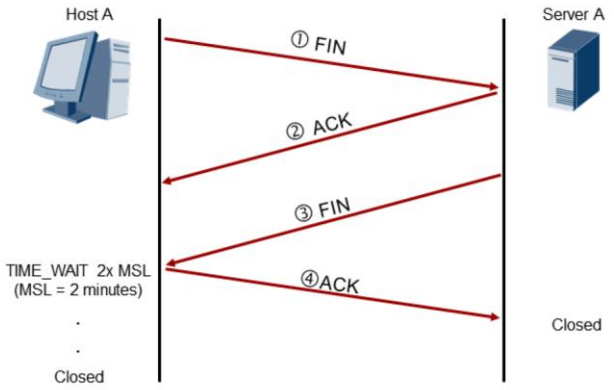


The TCP window field provides a means of flow control that governs the amount of data sent by the sender. This is achieved by returning a "window" with every TCP segment for which the ACK field is set, indicating a range of acceptable sequence numbers beyond the last segment successfully received. The window indicates the permitted number of octets that the sender may transmit before receiving further permission.

In the example, TCP transmission from host A to server A contains the current window size for host A. The window size for server A is determined as part of the handshake, which based on the transmission can be assumed as 2048. Once data equivalent to the window size has been received, an acknowledgement will be returned, relative to the number of bytes received, plus one. Following this, host A will proceed to transmit the next batch of data.

A TCP window size of 0 will effectively deny processing of segments, with exception to segments where the ACK, RST and URG code bits are set for incoming segments. Where a window size of 0 exists, the sender must still periodically check the window size status of receiving TCP to ensure any change in the window size is effectively reported, the period for retransmission is generally two minutes. When a sender sends periodic segments, the receiving TCP must still acknowledge with a sequence number announcement of the current window size of 0.

TCP Connection Termination



- Host A will ensure ACK is received by Server A before closing.

As part of the TCP connection termination process, a number of states are defined through which TCP will transition. These states include FIN-WAIT-1 that represents waiting for a connection termination (FIN) request from the remote TCP, or an acknowledgment of a connection termination request that was previously sent. The FIN-WAIT-2 represents waiting for a connection termination request from the remote TCP following which will generally transition to a TIME-WAIT state. A CLOSE-WAIT state indicates waiting for a locally defined connection termination request, typically when a servers application is in the process of closing.

The LAST-ACK state represents waiting for an acknowledgment of the connection termination request previously sent to the remote TCP (which includes an acknowledgment of its connection termination request). Finally, a TIME-WAIT state occurs and waits for enough time to pass to ensure that the remote TCP received the acknowledgment of its connection termination request. This period is managed by the Max Segment Lifetime (MSL) timer that defines a waiting period of 2 minutes. Following a wait period equal to two times the MSL, the TCP connection is considered closed/terminated.

User Datagram Protocol

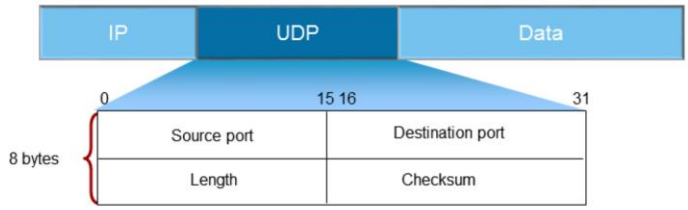
- UDP based data is sent without establishing a connection.

HUAWEI

The User Datagram Protocol, or UDP, represents an alternative to TCP and applied where TCP is found to act as an inefficient transport mechanism, primarily in the case of highly delay sensitive traffic. Where TCP is considered a segment, the UDP is recognized as a datagram form of Protocol Data Unit (PDU), for which a datagram can be understood to be a self-contained, independent entity of data carrying sufficient information to be routed from the source to the destination end system without reliance on earlier exchanges between this source and destination end systems and the transporting network, as defined in RFC 1594. In effect this means that UDP traffic does not require the establishment of a connection prior to the sending of data.

The simplified structure and operation of UDP makes it ideal for application programs to send messages to other programs, with a minimum of protocol mechanism such in the case of acknowledgements and windowing for example, as found in TCP segments. In balance however, UDP does not guarantee delivery of data transmission, nor protection from datagram duplication.

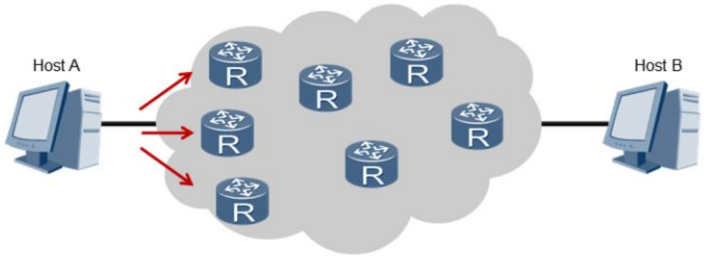
UDP Datagram Format



- UDP achieves minimal overhead for each datagram.
- Datagram delivery is not guaranteed with UDP.

The UDP header provides a minimalistic approach to the transport layer, implementing only a basic construct to help identify the destination port to which the UDP based traffic is destined, as well as a length field and a checksum value that ensures the integrity of the UDP header. In addition the minimal overhead acts as an ideal means for enabling more data to be carried per packet, favoring real time traffic such as voice and video communications where TCP provides a 20 byte overhead and mechanisms that influence delays, such as in the case of acknowledgements, however the lack of such fields means that datagram delivery is not guaranteed.

UDP Forwarding Behavior



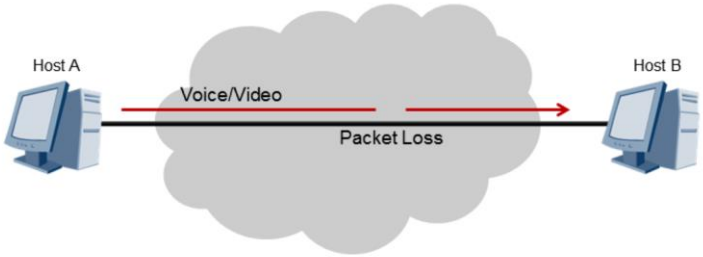
- UDP susceptible to the possibility of datagram duplication or non-orderly delivery of datagrams.

HUAWEI

Since UDP datagram transmission is not sent as a data stream, transmission of data is susceptible to datagram duplication. In addition, the lack of sequence numbers within UDP means that delivery of data transmission over various paths is likely to be received at the destination in an incorrect, non-sequenced order.

Where stream data is transported over UDP such as in the case of voice and video applications, additional protocol mechanisms may be applied to enhance the capability of UDP, as in the case of the real time transport protocol (RTP) which helps to support the inability of UDP by providing a sequencing mechanism using timestamps to maintain the order of such audio/video data streams, effectively supporting partial connection oriented behavior over a connectionless transport protocol.

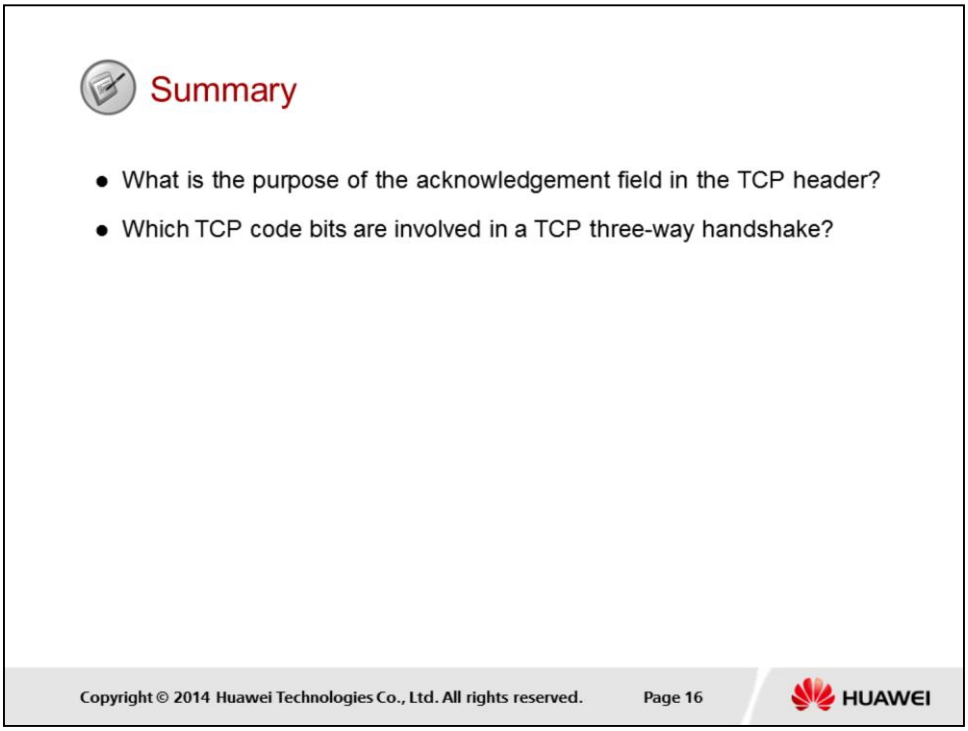
UDP Forwarding Behavior



- There are no acknowledgements, therefore lost packets are not retransmitted, this however is beneficial to delay sensitive data.

The general UDP forwarding behavior is highly beneficial to delay sensitive traffic such as voice and video. It should be understood that where a connection-oriented transport protocol is concerned, lost data would require replication following a delay period, during which time an acknowledgement by the sender is expected. Should the acknowledgement not be received, the data shall be retransmitted.

For delay sensitive data streams, this would result in incomprehensible audio and video transmissions due to both delay and duplication, as a result of retransmission from the point where acknowledgements are generated. In such cases, minimal loss of the data stream is preferable over retransmission, and as such UDP is selected as the transport mechanism, in support of delay sensitive traffic.



1. The acknowledgement field in the TCP header confirms receipt of the segment received by the TCP process at the destination. The sequence number in the TCP header of the received IP datagram is taken and incremented by 1. This value becomes the acknowledgement number in the returned TCP header and is used to confirm receipt of all data, before being forwarded along with the ACK code bit set to 1, to the original sender.
2. The three-way handshake involves SYN and ACK code bits in order to establish and confirm the connection between the two end systems, between which transmission of datagrams is to occur.

Thank you

www.huawei.com

Data Forwarding Scenario

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

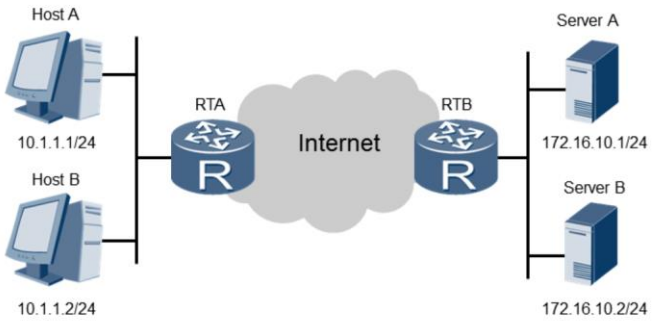
The TCP/IP protocol suite operates as a collection of rules in order to support the end-to-end forwarding of data, together with lower layer protocols such as those defined in the IEEE 802 standards. The knowledge of the lifecycle of data forwarding enables a deeper understanding of the IP network behavior for effective analysis of network operation and troubleshooting of networking faults. The entire encapsulation and decapsulation process therefore represents a fundamental part of all TCP/IP knowledge.

Objectives

Upon completion of this section, trainees will be able to:

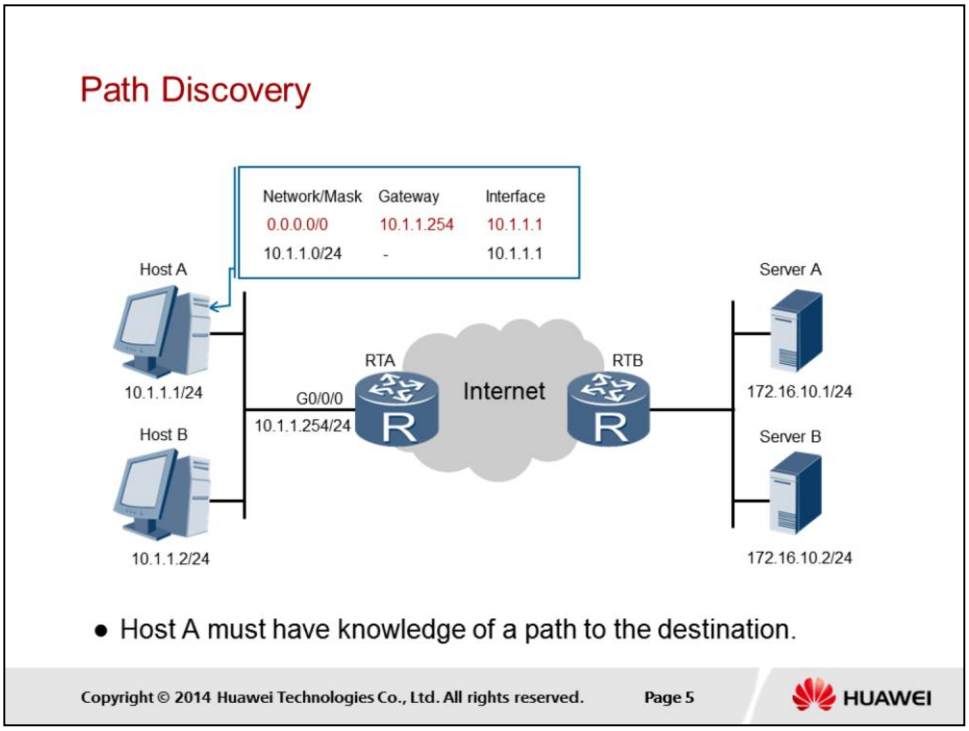
- Explain the process steps for data encapsulation and decapsulation.
- Troubleshoot basic data forwarding issues.

Scenario Introduction



- Data forwarding may be local or remote, however the general forwarding process is the same.

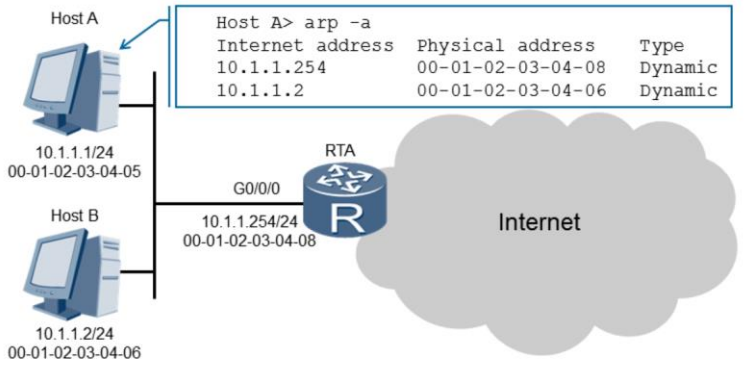
Data forwarding can be collectively defined as either local or remote for which the forwarding process relies on the application of the protocol stack in order to achieve end-to-end transmission. End systems may be part of the same network, or located in different networks, however the general forwarding principle to enable transmission between hosts follows a clear set of protocols that have been introduced as part of the unit. How these protocols work together shall be reinforced, as well as building the relationship between the upper layer TCP/IP protocols and the lower link layer based Ethernet protocol standards.



An end system that intends to forward data to a given destination must initially determine whether or not it is possible to reach the intended destination. In order to achieve this, the end system must go through a process of path discovery. An end system should be understood to be capable of supporting operation at all layers since its primary function is as a host to applications. In relation to this, it must also be capable of supporting lower layer operations such as routing and link layer forwarding (switching) in order to be capable of upper/application layer data forwarding. The end system therefore contains a table that represents network layer reachability to the network for which the upper layer data is destined.

End systems will commonly be aware of the network to which they reside, but may be without a forwarding path in cases where remote network discovery has not been achieved. In the example given, host A is in possession of a path to the destined network through the ‘any network’ address that was briefly introduced as part of the IP Addressing section. The forwarding table identifies that traffic should be forwarded to the gateway as a next hop via the interface associated with the logical address of 10.1.1.1.

ARP

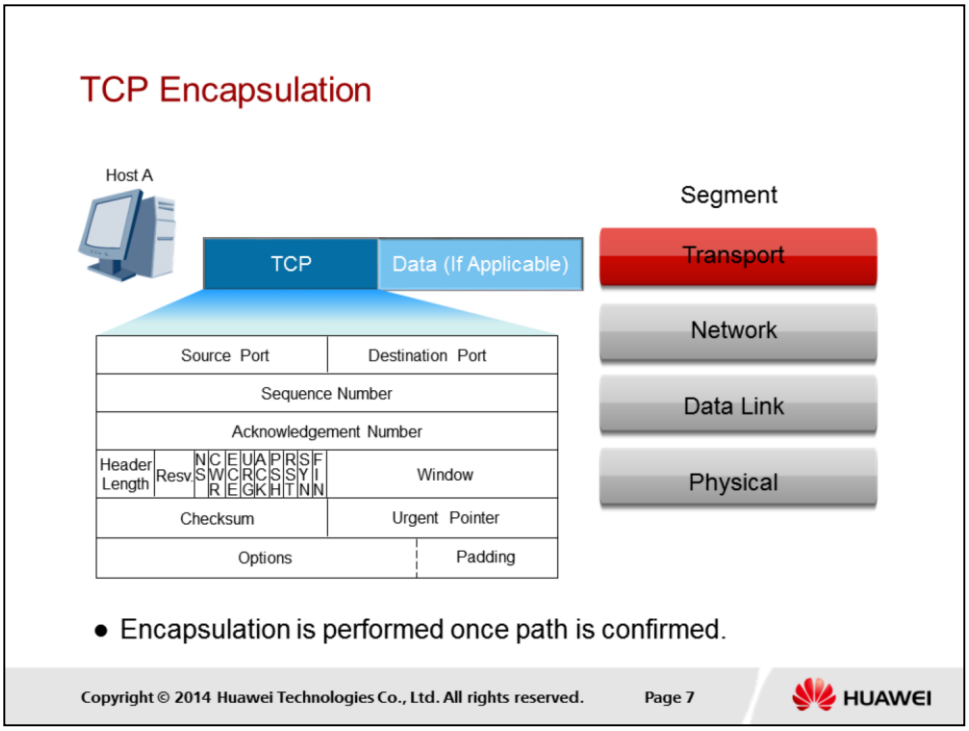


- The ARP cache table is used to discover the data link next hop.
- An unknown next hop will generate an ARP request.

HUAWEI

Following discovery of a feasible route towards the intended destination network, a physical next hop must also be discovered to facilitate frame forwarding. The TCP/IP protocol suite is responsible for determining this before packet encapsulation can proceed. The initial step involves determining whether a physical path exists to the next hop identified as part of the path discovery process.

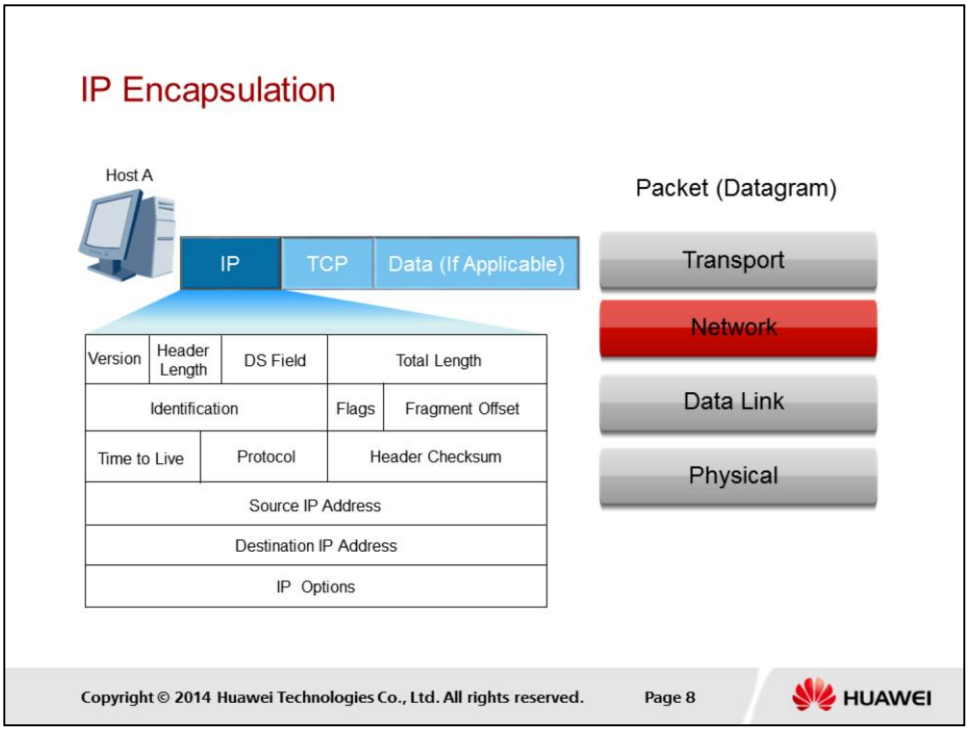
This requires that the ARP cache table be consulted to identify whether an association between the intended next hop and the physical path is known. From the example it can be seen that an entry to the next hop gateway address is present in the ARP cache table. Where an entry cannot be found, the Address Resolution Protocol (ARP) must be initiated to perform the discovery and resolve the physical path.



When both the logical and physical path forwarding discovery is complete, it is possible for encapsulation of data to be performed for successful transmission over IP/Ethernet based networks. Upper layer processes in terms of encryption and compression may be performed following which transport layer encapsulation will occur, identifying the source and destination ports via which upper layer data should be forwarded.

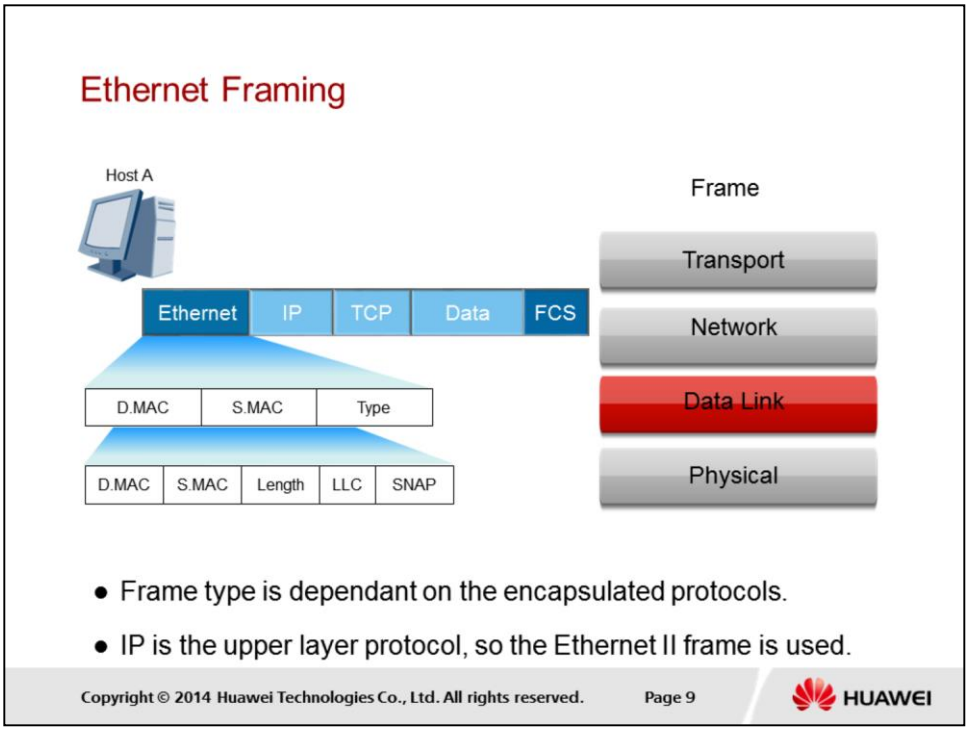
In the case of TCP, sequence and acknowledgement fields will be populated, code bits set as necessary with the ACK bit commonly applied. The window field will be populated with the current supported window size, to which the host will notify of the maximum data buffer that can be supported before data is acknowledged.

Values representing the TCP fields are included as part of the checksum, which is calculated using a ones compliment calculation process, to ensure TCP segment integrity is maintained once the TCP header is received and processed at the ultimate destination. In the case of basic TCP code operations, upper layer data may not always be carried in the segment, as in the case of connection synchronization, and acknowledgements to received data.



Following transport layer encapsulation, it is generally required that instructions be provided, detailing how transmission over one or more networks is to be achieved. This involves listing the IP source as well as the ultimate destination for which the packet is intended. IP packets are generally limited to a size of 1500 bytes by Ethernet, inclusive of the network and transport layer headers as well as any upper layer data.Initial packet size will be determined by Ethernet as the maximum transmission unit, or MTU to which packets will conform, therefore fragmentation will not occur at the source.

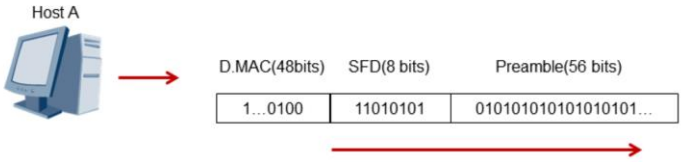
In the case that the MTU changes along the forwarding path, only then will fragmentation will be performed. The time to live field will be populated with a set value depending on the system, in ARG3 series routers, this is set with an initial value of 255. The protocol field is populated based on the protocol encapsulated prior to IP. In this case the protocol in question is TCP for which the IP header will populate the protocol field with a value of 0x06 as instruction for next header processing. Source and destination IP addressing will reflect the originating source and the ultimate destination.



Link layer encapsulation relies on IEEE 802.3 Ethernet standards for physical transmission of upper layer data over Ethernet networks. Encapsulation at the lower layers is performed by initially determining the frame type that is used.

Where the upper layer protocol is represented by a type value greater than 1536 (0x0600) as is the case with IP (0x0800), the Ethernet II frame type is adopted. The type field of the Ethernet II frame header is populated with the type value of 0x0800 to reflect that the next protocol to be processed following frame processing will be IP. The destination MAC address determines the next physical hop, which in this case represents the network gateway.

Frame Forwarding

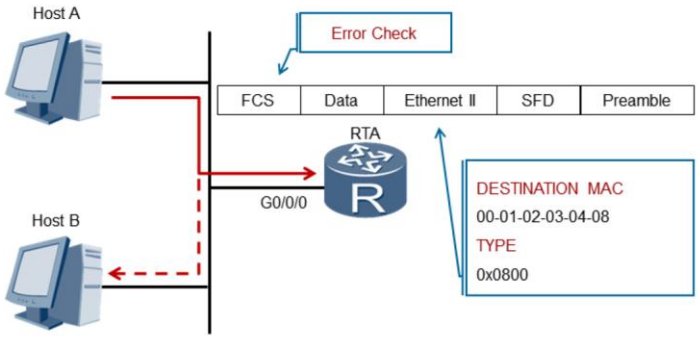


- Data link layer uses carrier sense to detect for existing traffic.
- Preamble and SFD used to synchronize with forwarded frame.

As part of the link layer operation, it is imperative to ensure that the transmission medium is clear of signals in shared collision domain. The host will first listen for any traffic on the network as part of CSMA/CD and should the line remain clear, will prepare to transmit the data. It is necessary for the receiving physical interface to be made aware of the incoming frame, so as to avoid loss of initial bit values that would render initial frames as incomplete. Frames are therefore preceded by a 64 bit value indicating to the link layer destination of the frame's imminent arrival.

The initial 56 bits represent an alternating 1, 0 pattern is called the preamble, and is followed immediately by an octet understood as the Start of Frame Delimiter (SFD). The final two bits of the SFD deviate from an alternating pattern to a 1,1 bit combination that notifies that the bits that follow represent the first bit values of the destination MAC address and therefore the start of the frame header.

Frame Processing



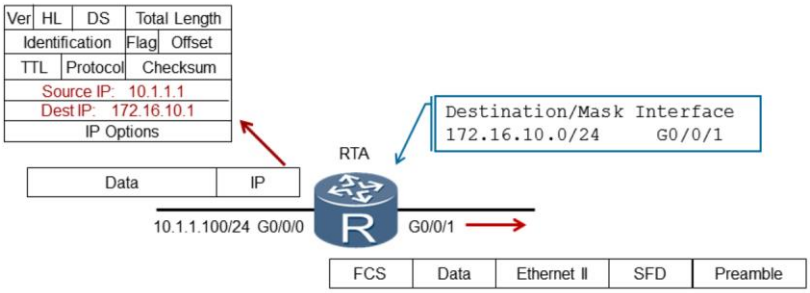
- Frame will be received by all in the same collision domain.
- Only the gateway (RTA) will process the frame.

As a frame is received by the link layer destination, it must go through a number of checks to determine its integrity as well as validity. If the frame was transmitted over a shared Ethernet network, other end stations may also receive an instance of the frame transmitted, however since the frame destination MAC address is different from the MAC address of the end station, the frame will be discarded.

Frames received at the intended destination will perform error checking by calculating the ones compliment value based on the current frame fields and compare this against the value in the Frame Check Sequence (FCS) field. If the values do not match, the frame will be discarded. Receiving intermediate and end systems that receive valid frames will need to determine whether the frame is intended for their physical interface by comparing the destination MAC address with the MAC address of the interface (or device in some cases).

If there is a match, the frame is processed and the type field is used to determine the next header to be processed. Once the next header is determined, the frame header and trailer are discarded.

Packet Processing



- Destination IP is checked against the address of the gateway.
- A new frame header is constructed following discovery process.

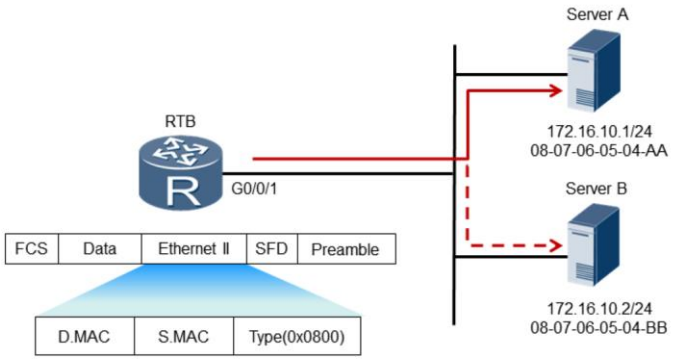
HUAWEI

The packet is received by the network layer, and in particular IP, at which point the IP header is processed. A checksum value exists at each layer of the protocol stack to maintain the integrity at all layers for all protocols. The destination IP is used to determine whether the packet has reached it's ultimate destination. The gateway however determines that this is not the case since the destination IP and the IP belonging to the gateway do not match.

The gateway must therefore determine the course of action to take with regards to routing the packet to an alternate interface, and forward the packet towards the network for which it is intended. The gateway must firstly however ensure that the TTL value has not reached 0, and that the size of the packet does not exceed the maximum transmission unit value for the gateway. In the event that the packet is larger than the MTU value of the gateway, fragmentation will generally commence.

Once a packet's destination has been located in the forwarding table of the gateway, the packet will be encapsulated in a new frame header consisting of new source and destination MAC addresses for the link layer segment, over which the resulting frame is to be forwarded, before being once again transmitted to the next physical hop. Where the next physical hop is not known, ARP will again be used to resolve the MAC address.

Frame Decapsulation

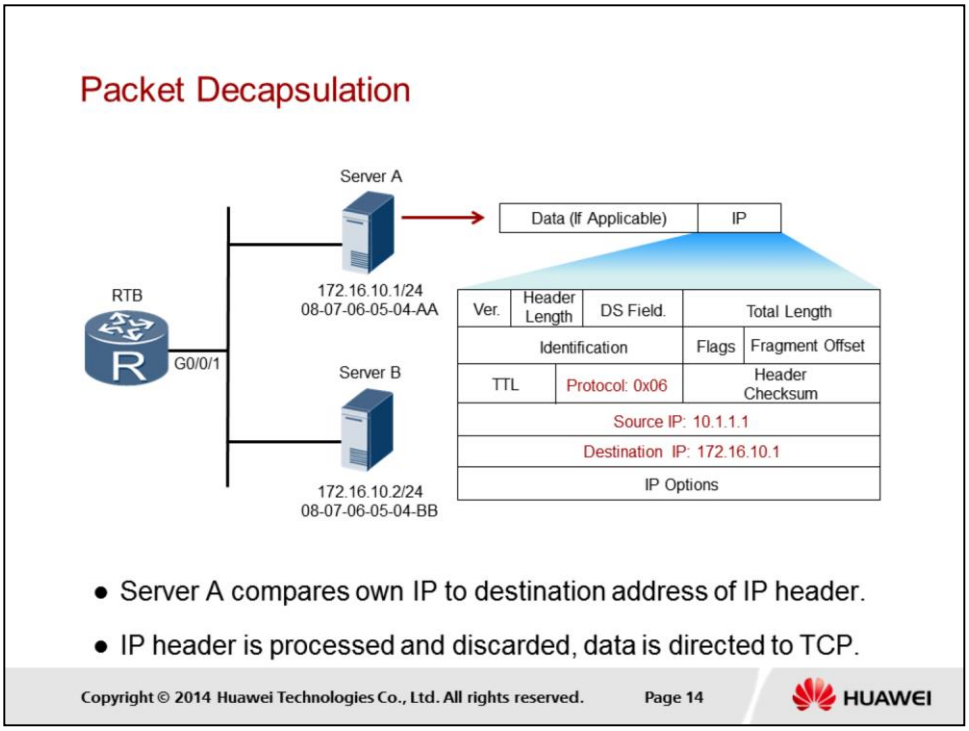


- Frame is forwarded with destination MAC address of Server A.
- Server A compares interface MAC to frame destination MAC.

HUAWEI

Frames received at the ultimate destination will initially determine whether the frame has arrived at the intended location. The example shows two servers on a shared Ethernet network over which both receive a copy of the frame.

The frame is ultimately discarded by server B since the destination MAC value and the interface MAC address of server B do not match. Server A however successfully receives the frame and learns that the MAC fields are the same, the integrity of the frame based on the FCS can also be understood to be correct. The frame will use the type field to identify 0x0800 as the next header, following which the frame header and trailer are discarded and the packet is received by IP.

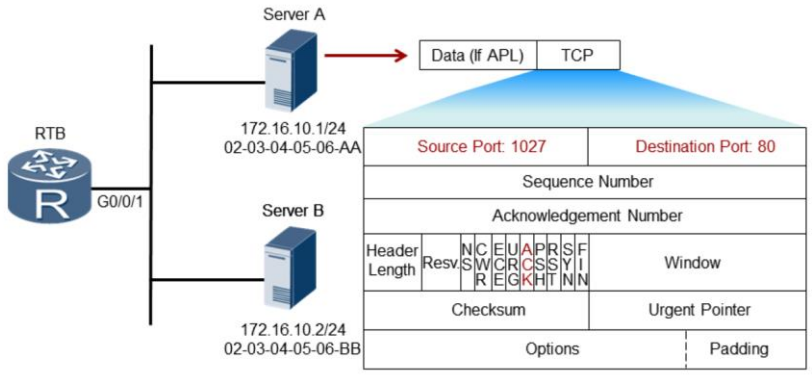


Upon reaching the ultimate destination, the IP packet header must facilitate a number of processes. The first includes validating the integrity of the packet header through the checksum field, again applying a ones compliment value comparison based on a sum of the IP header fields. Where correct, the IP header will be used to determine whether the destination IP matches the IP address of the current end station, which in this case is true.

If any fragmentation occurred during transmission between the source and the destination, the packet must be reassembled at this point. The identification field will collect the fragments belonging to a single data source together, the offset will determine the order and the flags field will specify when the reassembly should commence, since all fragments must be received firstly and a fragment with a flag of 0 will be recognized as the last fragment to be received.

A timer will then proceed during which time the reassembly must be completed, should reassembly fail in this time period, all fragments will be discarded. The protocol field will be used to identify the next header for processing and the packet header will be discarded. It should be noted that the next header may not always be a transport layer header, a clear example of where this can be understood is in the case of ICMP, which is understood to also be a network layer protocol with a protocol field value of 0x01.

Segment Decapsulation



- TCP header builds connection with the service at port 80.
- Parameters within the TCP header used to manage connection.

HUAWEI

In the case where a packet header is discarded, the resulting segment or datagram is passed to the transport layer for application-to-application based processing. The header information is received in this case by TCP (0x06).

In the example it can be understood that a TCP connection has already been established and the segment represents an acknowledgement for the transmission of HTTP traffic from the HTTP server to the acknowledging host. The host is represented by the port 1027 as a means to distinguish between multiple HTTP connections that may exist between the same source host and destination server. In receiving this acknowledgement, the HTTP server will continue to forward to the host within the boundaries of the window size of the host.

Summary

- What information is required before data can be encapsulated?
- What happens when a frame is forwarded to a destination to which it is not intended?
- How does the data in the frame ultimately reach the application it is intended for?
- When multiple sessions of the same application are active (e.g. multiple web browsers), how does the return data reach the correct session?

1. Prior to the encapsulation and forwarding of data, a source must have knowledge of the IP destination or an equivalent forwarding address such as a default address to which data can be forwarded. Additionally it is necessary that the forwarding address be associated with a physical next hop to which the data can be forwarded within the local network.
2. Any frame that is received by a gateway or end system (host) to which it is not intended, is subsequently dropped, following inspection of the destination MAC address in the frame header.
3. The delivery of data relies on the destination port number in the TCP and UDP headers to identify the application to which the data is intended. Following analysis of this value by the TCP or UDP protocol, the data is forwarded.
4. The source port of the TCP header for the HTTP traffic distinguishes between the different application sessions that are active. Return HTTP traffic from the HTTP server is able to identify each individual web browser session based on this source port number. For example, the source port of two separate requests for HTTP traffic originating from IP source 10.1.1.1 may originate from source ports 1028 and 1035, however the destination port in both cases remains as port 80, the HTTP server.

Thank you

www.huawei.com

Navigating The CLI

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

The implementation of Huawei devices in an enterprise network requires a level of knowledge and capability in the navigation of the VRP command line interface, and configuration of system settings. The principle command line architecture is therefore introduced as part of this section along with navigation, help functions and common system settings that are required to be understood for the successful configuration of any VRP managed device.

Objectives

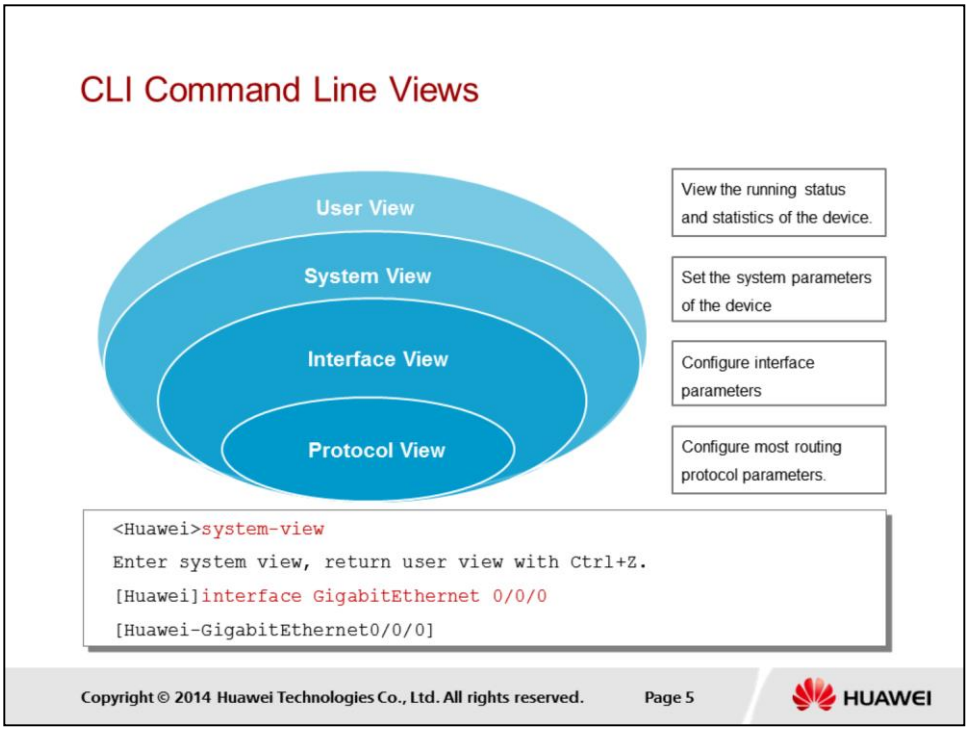
Upon completion of this section, trainees will be able to:

- Navigate the VRP command line interface.
- Configure basic VRP system settings.
- Perform basic VRP interface configuration and management.

Starting A Device

```
BIOS Creation Date : Jan  5 2013, 18:00:24
DDR DRAM init : OK
Start Memory Test ? ('t' or 'T' is test):skip
Copying Data : Done
Uncompressing : Done
……
Press Ctrl+B to break auto startup ... 1
Now boot from flash:/AR2220E-V200R007C00SPC600.cc,
……
<Huawei>
 Warning: Auto-Config is working. Before configuring the device, stop
Auto-Config. If you perform configurations when Auto-Config is
running, the DHCP, routing, DNS, and VTY configurations will be lost.
Do you want to stop Auto-Config? [y/n]:Y
```


The startup/boot process is the initial phase of operation for any administrator or engineer accessing Huawei based products operating with VRP. The boot screen informs of the system startup operation procedures as well as the version of the VRP image that is that currently implemented on the device, along with the storage location from where it is loaded. Following the initial startup procedure, an option for auto-configuration of the initial system settings prompts for a response, for which the administrator can choose whether to follow the configuration steps, or manually configure the basic system parameters. The auto-configuration process can be terminated by selecting the yes option at the given prompt.



The hierarchical command structure of VRP defines a number of command views that govern the commands for which users are able to perform operations. The command line interface has multiple command views, of which common views have been introduced in the example. Each command is registered to run in one or more command views, and such commands can run only after entering the appropriate command view. The initial command view of VRP is the User View, which operates as an observation command view for observing parameter statuses and general statistical information. For application of changes to system parameters, users must enter the System View. A number of sub command levels can also be found, in the form of the interface and protocol views for example, where sub system level tasks can be performed.

The command line views can be determined based on the parenthesis, and information contained within these parenthesis. The presence of chevrons identifies that the user is currently in the User View, whereas square brackets show that a transition to the System View has occurred.

CLI Functions

Command	Function
CTRL+A	Moves the cursor to the beginning of the current line.
CTRL+C	Stops performing current functions.
CTRL+Z	Returns to the user view.
CTRL+]	Stops incoming connections or redirects the connections.

```
<Huawei>system-view
Enter system view, return user view with Ctrl+Z.
[Huawei]^z  //Ctrl+Z
<Huawei>
```


The example demonstrates a selection of common system defined shortcut keys that are widely used to simplify the navigation process within the VRP command line interface. Additional commands are as follows:

CTRL+B moves the cursor back one character.

CTRL+D deletes the character where the cursor is located.

CTRL+E moves the cursor to the end of the current line.

CTRL+F moves the cursor forward one character.

CTRL+H deletes the character on the left side of the cursor.

CTRL+N displays the next command in the historical command buffer.

CTRL+P displays the previous command in the historical command buffer.

CTRL+W deletes the word on the left side of the cursor.

CTRL+X deletes all the characters on the left side of the cursor.

CTRL+Y deletes all the characters on the right side of the cursor.

ESC+B moves the cursor one word back.

ESC+D deletes the word on the right side of the cursor.

ESC+F moves the cursor forward one word.

CLI Functions

Command	Function
Backspace	Deletes the character on the left of the cursor, and moves the cursor to the left.
← or Ctrl+B	Moves the cursor a single character space to the left.
→ or Ctrl+F	Moves the cursor a single character space to the right.
TAB	Completes any incomplete keyword that is entered.

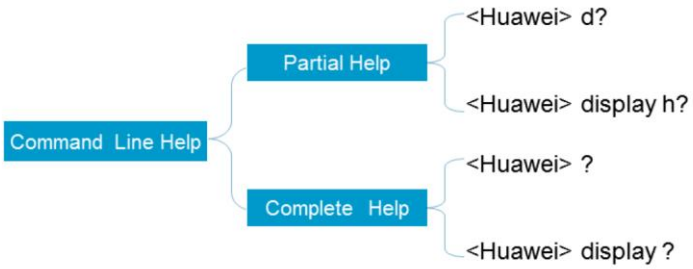
```
[Huawei]inter //TAB
[Huawei]interface
```

- The tab key will auto-complete an entered character string

Additional key functions can be used to perform similar operations, the backspace operation has the same behavior as using CTRL+H to delete a character to the left of the cursor. The left (←) and right (→) cursor keys can be used to perform the same operation as the CTRL+B and CTRL+F shortcut key functions. The down cursor key (↓) functions the same as Ctrl+N, and the up cursor key (↑) acts as an alternative to the CTRL+P operation.

Additionally, the command line functions support a means of auto completion where a command word is unique. The example demonstrates how the command word *interface* can be auto completed by partial completion of the word to such a point that the command is unique, followed by the tab key which will provide auto completion of the command word. Where the command word is not unique, the tab function will cycle through the possible completion options each time the tab key is pressed.

CLI Help Features




```
[Huawei]d?
  ddns                                dhcp
  dhcpv6                              diagnose
  display                             dns
  domain                              dot1x
```


There are two forms of help feature that can be found within the VRP, these come in the form of partial help and complete help functions. In entering a character string followed directly by a question mark (?), VRP will implement the partial help function to display all commands that begin with this character string. An example of this is demonstrated. In the case of the full help feature, a question mark (?) can be placed on the command line at any view to display all possible command names, along with descriptions for all commands pertaining to that view. Additionally the full help feature supports entry of a command followed by a question mark (?) that is separated by a space. All keywords associated with this command, as well as simple descriptions, are then displayed.

CLI Basic Device Setup

Command	Function
sysname	Configures the Device Name

```
<Huawei>system-view
Enter system view, return user view with Ctrl+Z.
[Huawei]sysname RTA
[RTA]
```

- The system name should be assigned to uniquely identify each device within an enterprise network.

For the majority of industries, it is likely that multiple devices will exist, each of which needs to be managed. As such, one of the first important tasks of device commissioning involves setting device names to uniquely identify each device in the network. The system name parameter on AR2200 series router is configured as Huawei by default, for the S5720 series of switch the default system name is HUAWEI. The implementation of the system name takes effect immediately after configuration is complete.

CLI Clock Settings

Command	Function
clock timezone	Sets the time zone.
clock datetime	Sets the current time and date.
clock daylight-saving-time	Sets the daylight saving time.

```
<Huawei>clock timezone BJ add 08:00:00
<Huawei>clock datetime 10:20:29 2016-04-11
<Huawei>display clock
2016-04-11 10:20:48
Thursday
Time Zone(BJ) : UTC+08:00
```


The system clock reflects the system timestamp, and is able to be configured to comply with the rules of any given region. The system clock must be correctly set to ensure synchronization with other devices and is calculated using the formula: Coordinated Universal Time (UTC) + Time zone offset + Daylight saving time offset. The clock datetime command is used to set the system clock following the *HH:MM:SS YYYY-MM-DD* formula. It should be noted however that if the time zone has not been configured or is set to 0, the date and time set are considered to be UTC, therefore it is recommended that the clock timezone be set firstly before configuring the system time and date.

The setting of the local timezone is achieved using the clock timezone command and is implemented based on the *time-zone-name* { *add* | *minus* } *offset* formula, where the add value indicates that the time of *time-zone-name* is equal to the UTC time plus the time offset and minus indicates the time of *time-zone-name* is equal to the UTC time minus the time offset.

Certain regions require that the daylight saving time be implemented to maintain clock synchronization with any change in the clock timezone during specific periods of the year. VRP is able to support daylight saving features for both fixed dates and dates which are determined based on a set of predetermined rules. For example, daylight saving in the United Kingom occurs on the last Sunday of March and the last Sunday of October, therefore rules can be applied to ensure that changes occur based on such fluctuating dates.

CLI Header Messages

Command	Function
header login	Sets the header that is displayed on a terminal when a user is authenticated by a device
header shell	Sets the header that is displayed on a terminal after the user logs into the device.

```
[Huawei]header login information "welcome to huawei certification!"
[Huawei]header shell information "Please don't reboot the device!"
.....
welcome to huawei certification!
Login authentication
Password:
Please don't reboot the device!
<Huawei>
```


The header command provides a means for displaying notifications during the connection to a device. The login header indicates a header that is displayed when the terminal connection is activated, and the user is being authenticated by the device. The shell header indicates a header that is displayed when the session is set up, after the user logs in to the device. The header information can be applied either as a text string or retrieved from a specified file. Where a text string is used, a start and end character must be defined as a marker to identify the information string, where in the example the " character defines the information string. The string represents a value in the range of 1 to 2000 characters, including spaces. The information based header command follows the format of *header* { *login* | *shell* } information *text* where information represents the information string, including start and end markers.

In the case of a file based header, the format header { *login* | *shell* } file *file-name* is applied, where *file-name* represents the directory and file from which the information string can be retrieved.

CLI Command Levels

User Level	Command Level	Name
0	0	Visit level
1	0 and 1	Monitoring level
2	0,1 and 2	Configuration level
3-15	0,1,2 and 3	Management level

```
<Huawei> system-view
[Huawei]command-privilege level 3 view user save
```

- Privilege levels manage user access to commands.

The system structures access to command functions hierarchically to protect system security. The system administrator sets user access levels that grant specific users access to specific command levels. The command level of a user is a value ranging from 0 to 3, whilst the user access level is a value ranging from 0 to 15. Level 0 defines a visit level for which access to commands that run network diagnostic tools, (such as ping and traceroute), as well as commands such as telnet client connections, and select display commands.

The Monitoring level is defined at a user level of 1 for which command levels 0 and 1 can be applied, allowing for the majority of display commands to be used, with exception to display commands showing the current and saved configuration. A user level of 2 represents the Configuration level for which command levels up to 2 can be defined, enabling access to commands that configure network services provided directly to users, including routing and network layer commands. The final level is the Management level which represents a user level of 3 through to 15 and a command level of up to 3, enabling access to commands that control basic system operations and provide support for services.

These commands include file system, FTP, TFTP, configuration file switching, power supply control, backup board control, user management, level setting, system internal parameter setting, and debugging commands for fault diagnosis. The given example demonstrates how a command privilege can be changed, where in this case, the save command found under the user view requires a command level of 3 before the command can be used.

CLI User Interfaces

User Interface	Relative Number
Console	0
VTY	0-4

```
<Huawei>system-view
[Huawei]user-interface vty 0 4
[Huawei-ui-vty0-4]
```

- The VTY number can be extended to a range of 0-14 for additional Telnet/SSH user connections.

Each user interface is represented by a user interface view or command line view provided by the system. The command line view is used to configure and manage all the physical and logical interfaces in asynchronous mode. Users wishing to interface with a device will be required to specify certain parameters in order to allow a user interface to become accessible. Two common forms of user interface implemented are the console interface (CON) and the virtual teletype terminal (VTY) interface.

The console port is an asynchronous serial port provided by the main control board of the device, and uses a relative number of 0. VTY is a logical terminal line that allows a connection to be set up when a device uses telnet services to connect to a terminal for local or remote access to a device. A maximum of 15 users can use the VTY logical user interface to log in to the device by extending the range from 0 – 4 achieved by applying the *user-interface maximum-vty 15* command. If the set maximum number of login users is 0, no users are allowed to log in to the router through telnet or SSH. The *display user-interface* command can be used to display relevant information regarding the user interface.

CLI Terminal Attributes

Command	Function
idle-timeout	Sets the timeout duration of the user connection.
screen-length	Sets the number of lines displayed on each terminal screen after a command is executed.
history-command max-size	Sets the size of the history command buffer.

```
# Set the size of the history command buffer to 20.
<Huawei>system-view
[Huawei]user-interface console 0
[Huawei-ui-console0]history-command max-size 20
# Set the timeout duration to 1 minute and 30 seconds.
[Huawei-ui-console0]idle-timeout 1 30
```


For both the console and VTY terminal interfaces, certain attributes can be applied to modify the behavior as a means of extending features and improving security. A user allows a connection to remain idle for a given period of time presents a security risk to the system. The system will wait for a timeout period before automatically terminating the connection. This idle timeout period on the user interface is set to 10 minutes by default .

Where it may be necessary to increase or reduce the number of lines displayed on the screen of a terminal when using the *display* command for example, the screen-length command can be applied. This by default is set to 24 however is capable of being increased to a maximum of 512 lines. A screen-length of 0 however is not recommended since no output will be displayed.

For each command that is used, a record is stored in the history command buffer which can be retrieved through navigation using the (↑) or CTRL+P and the (↓) or Ctrl+N key functions. The number of recorded commands in the history command buffer can be increased using the *history-command max-size* command to define up to 256 stored commands. The number of commands stored by default is 10.

CLI Interface Permissions

Command	Function
user privilege	Configures the user level.
set authentication password	Configures a local authentication password.

```
# Set the user level on the VTY0 user interface to 2.
<Huawei>system-view
[Huawei]user-interface vty 0
[Huawei-ui-vty0]user privilege level 2
[Huawei-ui-vty0-4]set authentication password cipher
 Enter Password(<8-128>):huawei123
```

  HUAWEI

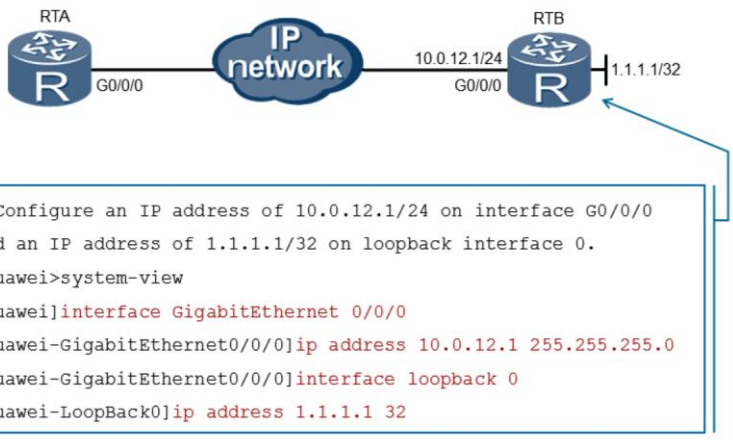
Access to user terminal interfaces provides a clear point of entry for unauthorized users to access a device and implement configuration changes. As such the capability to restrict access and limit what actions can be performed is necessary as a means of device security. The configuration of user privilege and authentication are two means by which terminal security can be improved. User privilege allows a user level to be defined which restricts the capability of the user to a specific command range. The user level can be any value in the range of 0 – 15, where values represent a visit level (0), monitoring level (1), configuration level (2), and management level (3) respectfully.

Authentication restricts a users capability to access a terminal interface by requesting the user be authenticated using a password or a combination of username and password before access via the user interface is granted. In the case of VTY connections, all users must be authenticated before access is possible. For all user interfaces, three possible authentication modes exist, in the form of AAA, password authentication and non-authentication. AAA provides user authentication with high security for which a user name and password must be entered for login. Password authentication requires that only the login password is needed therefore a single password can be applied to all users. The use of non-authentication removes any authentication applied to a user interface. It should be noted that the console interface by default uses the non-authentication mode.

It is generally recommended that for each user that is granted telnet access, the user be identified through usernames and passwords to allow for distinction of individual users. Each user should also be granted privilege

levels, based on each users role and responsibility.

CLI Interface Configuration



HUAWEI

In order to run IP services on an interface, an IP address must be configured for the interface. Generally, an interface needs only the primary IP address. In special cases, it is possible for a secondary IP address to be configured for the interface. For example, when an interface of a router such as the AR2200 connects to a physical network, and hosts on this physical network belong to two network segments.

In order to allow the AR2200 to communicate with all the hosts on the physical network, configure a primary IP address and a secondary IP address for the interface. The interface has only one primary IP address. If a new primary IP address is configured on an interface that already has a primary IP address, the new IP address overrides the original one. The IP address can be configured for an interface using the command ip address <*ip-address* > { mask | *mask-length* } where *mask* represents the 32 bit subnet mask e.g. 255.255.255.0, and *mask-length* represents the alternative mask-length value e.g. 24, both of which can be used interchangeably.

The loopback interface represents a logical interface that is applied to represent a network or IP host address, and is often used as a form of management interface in support of a number of protocols through which communication is made to the IP address of the loopback interface, as opposed to the IP address of the physical interface on which data is being received.

Summary

- How many users are able to connect via the console interface at any given time?
- What is the state of the loopback interface 0 when the command *loopback interface 0* is used?

1. The console interface is capable of supporting only a single user at any given time; this is represented by the console 0 user interface view.
2. The loopback interface represents a logical interface that is not present in a router until it is created. Once created, the loopback interface is considered up. On ARG3 devices, the loopback interfaces can however be shut down.

Thank you

www.huawei.com

File System Navigation and Management

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

The file system represents the underlying platform on which VRP operates, and where system files are stored within the physical storage devices of the product. The capability to navigate and manage this file system is necessary to ensure effective management of the configuration files, VRP software upgrades and ensure that the physical devices contained within each product are well maintained.

Objectives

Upon completion of this section, trainees will be able to:

- Successfully navigate the device file system
- Manipulate the file system files and folders.
- Manage Huawei router and switch storage devices.

Viewing the File System

Function	Command
Change directory	cd
View current directory	pwd
View content of directory	dir
View file content	more

```
<Quidway>dir
Directory of flash:/
  Idx   Attr     Size(Byte)  Date        Time        FileName
    0   drw-           -     Apr 10 2016 09:30:35    src
    1   -rw-          28     Apr 10 2016 09:31:38    private-data.txt
    2   -rw-         120     Apr 10 2016 09:32:38    wzbk1.cfg
32,004 KB total (31,995 KB free)
```


The file system manages files and directories on the storage devices. It can create, delete, modify, or rename a file or directory, or display the contents of a file.

The file system has two functions: managing storage devices and managing the files that are stored on those devices. A number of directories are defined within which files are stored in a logical hierarchy. These files and directories can be managed through a number of functions which allow the changing or displaying of directories, displaying files within such directories or sub-directories, and the creation or deletion of directories.

Common examples of file system commands for general navigation include the *cd* command used to change the current directory, *pwd* to view the current directory and *dir* to display the contents of a directory as shown in the example. Access to the file system is achieved from the User View.

Manipulating the File System

Function	Command
Make directory	mkdir
Remove directory	rmdir

```
<Quidway>mkdir test
Info: Create directory flash:/test......Done.
<Quidway>dir
Directory of flash:/
  Idx   Attr     Size(Byte)  Date       Time        FileName
    0   drw-          -      Apr 10 2016 09:30:35   src
    1   -rw-         28      Apr 10 2016 09:31:38   private-data.txt
    2   -rw-        120      Apr 10 2016 09:32:38   wzbk1.cfg
    3   drw-          -      Apr 10 2016 09:53:11   test
32,004 KB total (31,995 KB free)
```


Making changes to the existing file system directories generally relates to the capability to create and delete existing directories within the file system. Two common commands that are used in this case. The *mkdir directory* command is used to create a folder in a specified directory on a designated storage device, where *directory* refers to the name given to the directory and for which the directory name can be a string of 1 to 64 characters. In order to delete a folder within the file system, the *rmdir directory* command is used, with *directory* again referring to the name of the directory. It should be noted that a directory can only be deleted if there are no files contained within that directory.

Manipulating the File System

Function	Command
Copy file	copy
Move file	move
Rename file	rename

```
<Quidway>rename test huawei
Rename flash:/test to flash:/huawei ?[Y/N]:y
Info: Rename file flash:/test to flash:/huawei ......Done.
<Quidway>dir
Directory of flash:/
  Idx   Attr     Size(Byte)  Date        Time        FileName
    0   drw-          -      Apr 10 2016 09:30:35    src
    1   -rw-         28      Apr 10 2016 09:31:38    private-data.txt
    2   -rw-         120     Apr 10 2016 09:32:38    wzbk1.cfg
    3   drw-          -      Apr 10 2016 09:53:11    huawei

32,004 KB total (31,995 KB free)
```


Making changes to the files within a file system includes copying, moving, renaming, compressing, deleting, undeleting, deleting files in the recycle bin, running files in batch and configuring prompt modes. Creating a duplicate of an existing file can be done using the copy source-filename destination-filename command, where if the destination-filename is the same as that of an existing file (source-filename), the system will display a message indicating that the existing file will be replaced. A target file name cannot be the same as that of a startup file, otherwise the system displays a message indicating that the operation is invalid and that the file is a startup file.

The move source-filename destination-filename command can be used to move files to another directory. After the move command has been successfully executed, the original file is cut and moved to the defined destination file. It should be noted however that the move command can only move files in the same storage device.

Manipulating the File System

Function	Command
Delete or permanently delete file	delete /unreserved
Recover file	undelete
Permanently clear the recycle bin	reset recycle-bin

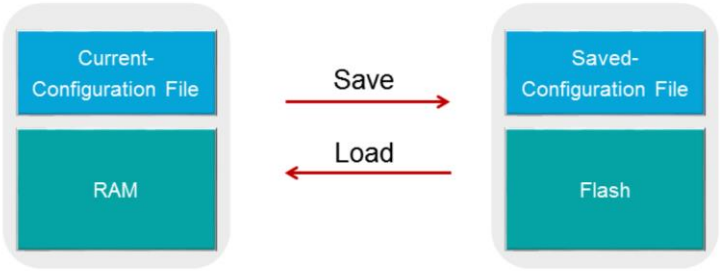
```
<Quidway>delete /unreserved flash:/wzbk1.cfg
<Quidway>dir
Directory of flash:/
  Idx   Attr     Size(Byte)  Date        Time        FileName
    0   drw-           -     Apr 10 2016 09:30:35    src
    1   -rw-          28     Apr 10 2016 09:31:38    private-data.txt
    2   drw-           -     Apr 10 2016 09:53:11    huawei

32,004 KB total (30,995 KB free)
```


For the removal of files within a file system, the delete function can be applied using the command *delete* [*/unreserved*] [*/force*] { *filename* | *device-name* }. Generally files that are deleted are directed to a recycle bin from where files can recovered using the *undelete* { *filename* | *device-name* } command, however should the /unreserved command be used, the file will be permanently deleted. The system will generally display a message asking for confirmation of file deletion, however if the */force* parameter is included, no prompt will be given. The *filename* parameter refers to the file which is to be deleted, while the *device-name* parameter defines the storage location.

Where a file is directed to the recycle bin, it is not permanently deleted and can be easily recovered. In order to ensure that such files in the recycle bin are deleted permanently, the *reset recycle-bin* [*filename*]command can be applied, where the *filename* parameter can be used to define a specific file for permanent deletion.

Configuration File Management System



- Current configuration loaded from saved configuration in system flash memory at system startup.

HUAWEI

When powered on, the device retrieves configuration files from a default save path to initialize itself, which is then stored within the RAM of the device. If configuration files do not exist in the default save path, the router uses default initialization parameters.

The current-configuration file indicates the configurations in effect on the device when it is actually running. When the configuration is saved, the current configuration is stored in a saved-configuration file within the storage location of the device. If the device loaded the current configuration file based on default initialization parameters, a saved-configuration file will not exist in the storage location of the default save path, but will be generated once the current configuration is saved.

Viewing Configuration Files

Command	Function
display current-configuration	View the current configuration
display saved-configuration	View the saved configuration

```
<Huawei>display current-configuration
#
sysname Huawei
.....
#
Return
<Huawei>display saved-configuration
#
sysname Huawei
.....
#
Return
```

  HUAWEI

Using the *display current-configuration* command, device parameters that take effect can be queried. If default values of certain parameters are being used, these parameters are not displayed. The current configuration command includes a number of parameters that allow for filtering of the command list during the used of the display function. The *display current-configuration | begin {regular-expression}* is an example of how the current configuration can be used to display active parameters that begin with a specific keywords or expressions. An alternative to this command is the *display current-configuration | include {regular-expression}* which allows parameters that include a specific keyword or experssion within the current-configuration file.

The *display saved-configuration* [*last* | *time*] shows the output of the stored configuration file used at startup to generate the current-configuration. Where the *last* parameter is used it displays the configuration file used in the current startup. The configuration file is displayed only when it is configured for the current startup. The *time* parameter will display the time when the configuration was last saved.

Saving the Configuration File

Command	Function
Save	Save the current configuration

```
<Huawei>save
The current configuration will be written to the device.
Are you sure to continue?[Y/N]y
It will take several minutes to save configuration file, please
  wait..............
  Configuration file had been saved successfully
  Note: The configuration file will take effect after being activated
```


Using the *save* [*configuration-file*] command will save the current configuration information to a default storage path. The *configuration-file* parameter allows the current configuration information to be saved to a specified file. Running the *save* command with the *configuration-file* parameter does not affect the current startup configuration file of the system. When *configuration-file* is the same as the configuration file stored in the default storage path of the system, the function of this command is the same as that of the *save* command.

The example demonstrates the use of the *save* command to save the current-configuration, which by default will be stored to the default vrpcfg.zip file in the default storage location of the device.

Viewing the Startup Parameters

Command	Function
Display startup	View the current startup parameters

```
<Huawei>display startup
MainBoard:
  Configured startup system software:     flash:/ar2220.cc
  Startup system software:                flash:/ar2220.cc
  Next startup system software:           NULL
  Startup saved-configuration file:       flash:/vrpcfg.zip
  Next startup saved-configuration file:  flash:/vrpcfg.zip
  Startup paf file:                       NULL
  Next startup paf file:                  NULL
  Startup license file:                   NULL
  Next startup license file:              NULL
  Startup patch package:                  NULL
  Next startup patch package:             NULL
```


The currently used save configuration file can be discovered through the use of the *display startup* command. In addition the *display startup* command can be used to query the name of the current system software file, name of the next system software file, name of the backup system software file, names of the four currently used (if used) system software files, and names of the next four system software files. The four system software files are the aforementioned configuration file, voice file, patch file, and license file.

Changing the Startup Parameters

Command	Function
startup saved-configuration	Specify saved configuration file to load at startup

```
<Huawei>startup saved-configuration flash:/huawei.zip
Info: Succeeded in setting the configuration for booting system.
<Huawei>display startup
MainBoard:
  Configured startup system software:     flash:/ar2220.cc
  Startup system software:                flash:/ar2220.cc
  Next startup system software:           NULL
  Startup saved-configuration file:       flash:/vrpcfg.zip
  Next startup saved-configuration file:  flash:/huawei.zip
  Startup paf file:                       NULL
  Next startup paf file:                  NULL
  Startup license file:                   NULL
  Next startup license file:              NULL
  Startup patch package:                  NULL
  Next startup patch package:             NULL
```


Following discovery of the startup saved-configuration file, it may be necessary to define a new configuration file to be loaded at the next startup. If a specific configuration file is not specified, the default configuration file will be loaded at the next startup.

The filename extension of the configuration file must be .cfg or .zip, and the file must be stored in the root directory of a storage device. When the router is powered on, it reads the configuration file from the flash memory by default to initialize. The data in this configuration file is the initial configuration. If no configuration file is saved in the flash memory, the router uses default parameters to initiate.

Through the use of the startup saved-configuration [*configuration-file*] where the configuration-file parameter is the configuration file to be used at startup, it is possible to define a new configuration file to initialize at the next system startup.

Comparing Configuration Files

Command	Function
compare configuration	Compare configuration files

```
<Huawei>compare configuration
====== Current configuration line 36 ======
 ip address 10.1.1.1 255.255.255.0
#
interface GigabitEthernet0/0/2
#
interface GigabitEthernet0/0/3
#
interface NULL0
 ====== Configuration file line 37 ======
interface GigabitEthernet0/0/2
#
interface GigabitEthernet0/0/3
#
interface NULL0
```


When the *compare configuration* [*configuration-file*] [*current-line-number save-line-number*] command is used, the system performs a line by line comparison of the saved configuration with the current configuration starting from the first line. If the *current-line-number save-line-number* parameters are specified, the system skips the non relevant configuration before the compared lines and continues to find differences between the configuration files.

The system will then proceed to output the configuration differences between the saved configuration and the current configuration files. The comparison output information is restricted to 150 characters by default. If the comparison requires less than 150 characters, all variations until the end of two files are displayed.

Clearing the Configuration File

Command	Function
reset saved-configuration	Erase saved configuration file

```
<Huawei>reset saved-configuration
Warning: This will delete the configuration in the flash memory.
The device configurations will be erased to reconfigure. Are you
soure? [Y/N]:y
Info: Clear the configuration in the device successfully.
```


The *reset saved-configuration* command is used in order to delete a device startup configuration file from the storage device. When performed, the system compares the configuration files used in the current startup and the next startup when deleting the configuration file from the router.

If the two configuration files are the same, they are deleted at the same time after this command is executed. The default configuration file is used when the router is started next time. If the two configuration files are different, the configuration file used in the current startup is deleted after this command is executed.

If no configuration file is configured for the device current startup, the system displays a message indicating that the configuration file does not exist after this command is executed. Once the reset saved-configuration command is used, a prompt will be given to confirm the action, for which the user is expected to confirm, as shown in the example.

Storage Device Types

- SDRAM
- Flash
- NVRAM
- SD Card
- USB

```
<Huawei>display version
......
SDRAM Memory Size     : 1024     M bytes
Flash Memory Size     : 512       M bytes
NVRAM Memory Size     : 512      K bytes
......
```


The storage devices are product dependant, and include flash memory, SD cards, or USB flash drives. The AR2200E router for example has a built-in flash memory and a built-in SD card (in slot sd1). The router provides two reserved USB slots (usb0 and usb1) and an SD card slot (sd0). For the S5700 it includes a built in flash memory with a capacity that varies dependant on the model, with 64MB supported in the S5700C-HI, S5700-LI, S5700S-LI and S5710-EI models, and 32 MB for all others. The details regarding the Huawei product storage devices can be detailed by using the *display version* command as shown.

Erasing Storage Devices

```
<Huawei>format flash:
All data(include configuration and system startup file) on flash:
will be lost, proceed with format? (y/n)[n]:

<Huawei>format sd1:
All data(include configuration and system startup file) on sd1: will
be lost, proceed with format? (y/n)[n]:
```

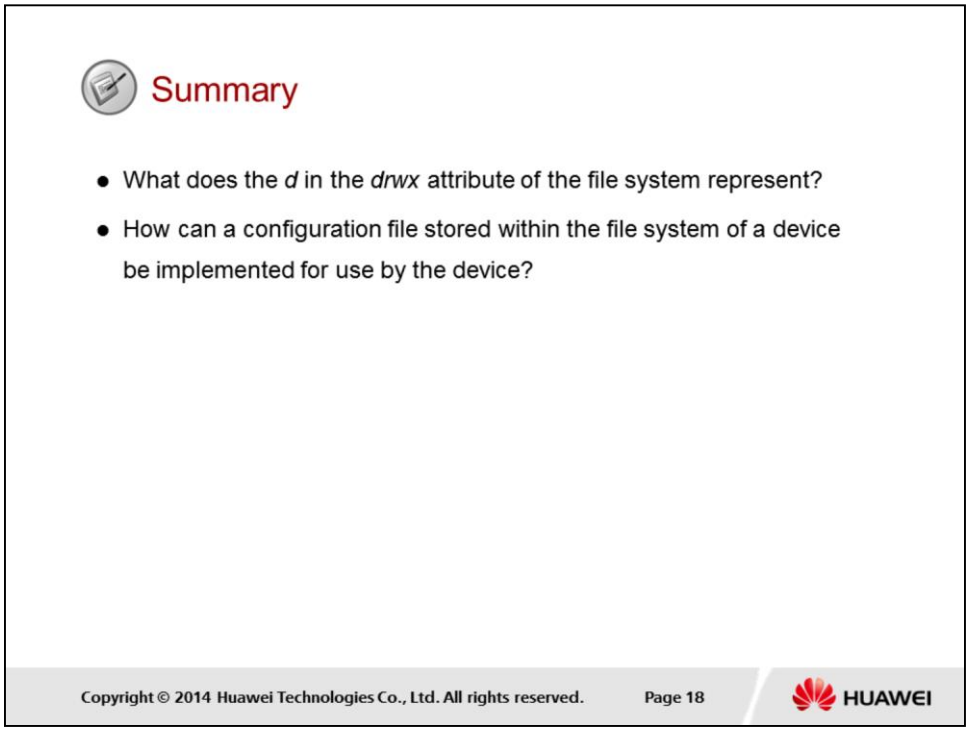
- Care should be taken when using the format commands, as data will be lost.

Formatting a storage device is likely to result in the loss of all files on the storage device, and the files cannot be restored, therefore extra care should be taken when performing any format command and should be avoided unless absolutely necessary. The format [*storage-device*] command is used along with the *storage-device* parameter to define the storage location which is required to be formatted.

Repairing the Storage Device

```
<Huawei>fixdisk flash:
Fixdisk flash: will take long time if needed
%Fixdisk flash: completed.
<Huawei>fixdisk sd1:
sd1:/  - disk check in progress.............sd1:/  - Volume is OK
total # of clusters:  481,869
# of free clusters:  455,777
# of bad clusters:
total free space:  1,780 Mb
.............. max contiguous free space:  1,789,952,000 bytes
# of files:  22
......
%Fixdisk sd1: completed.
```


When the terminal device displays that the system has failed, the *fixdisk* command can be used to attempt to fix the abnormal file system in the storage device, however it does not provide any guarantee as to whether the file system can be restored successfully. Since the command is used to rectify problems, if no problem has occurred in the system it is not recommended that this command be run. It should also be noted that this command does not rectify device-level problems.



1. The file system attribute *d* represents that the entry is a directory in the file system. It should be noted that this directory can only be deleted once any files contained within the directory have been deleted. The remaining rwx values refer to whether the directory (or file) can be read, written to, and/or executed.
2. A configuration may be saved under a separate name from the default vrpcfg.zip file name and stored within the storage device of the router or switch. If this file is required to be used as the active configuration file in the system, the command *startup saved-configuration <configuration-file-name>* should be used where the configuration-file-name refers to the file name and file extension.

Thank you

www.huawei.com

VRP Operating System Image Management

HUAWEI TECHNOLOGIES CO., LTD.

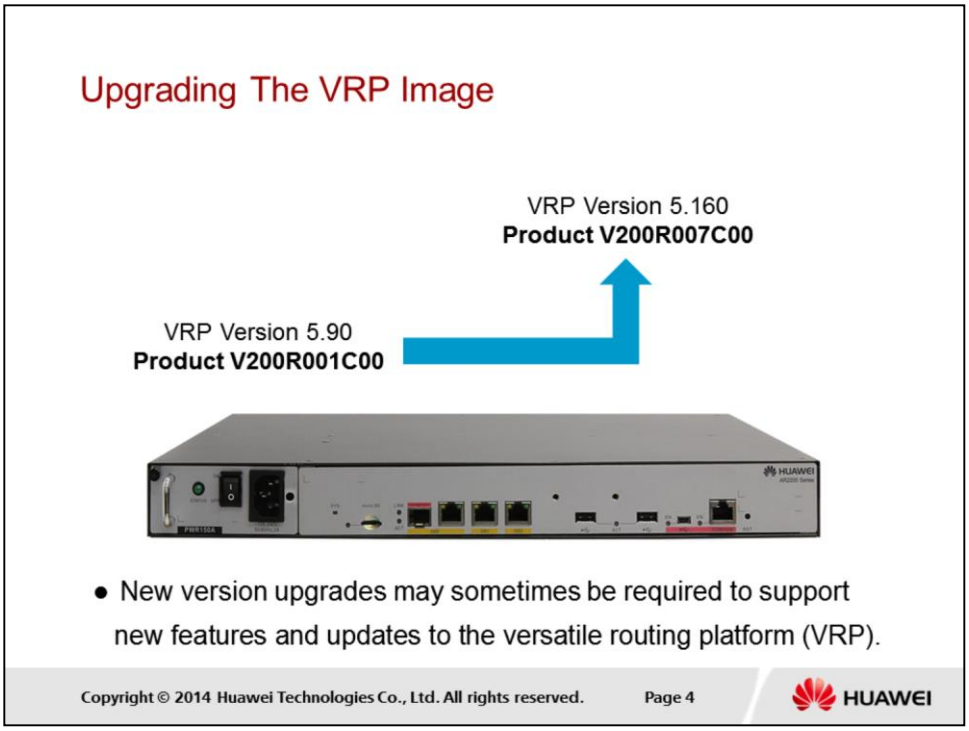
Foreword

Effective network administration and management within an enterprise network relies on all devices maintaining backup files in the event of system failures or other events that may result in loss of important systems files and data. Remote servers that use the file transfer protocol (FTP) service are often used to ensure files are maintained for backup and retrieval purposes as and when needed. The means for establishing communication with such application services is introduced in this section.

Objectives

Upon completion of this section, trainees will be able to:

- Explain the importance of maintaining up-to-date versions of VRP.
- Establish a client relationship with an FTP server.
- Successfully upgrade a VRP system image.



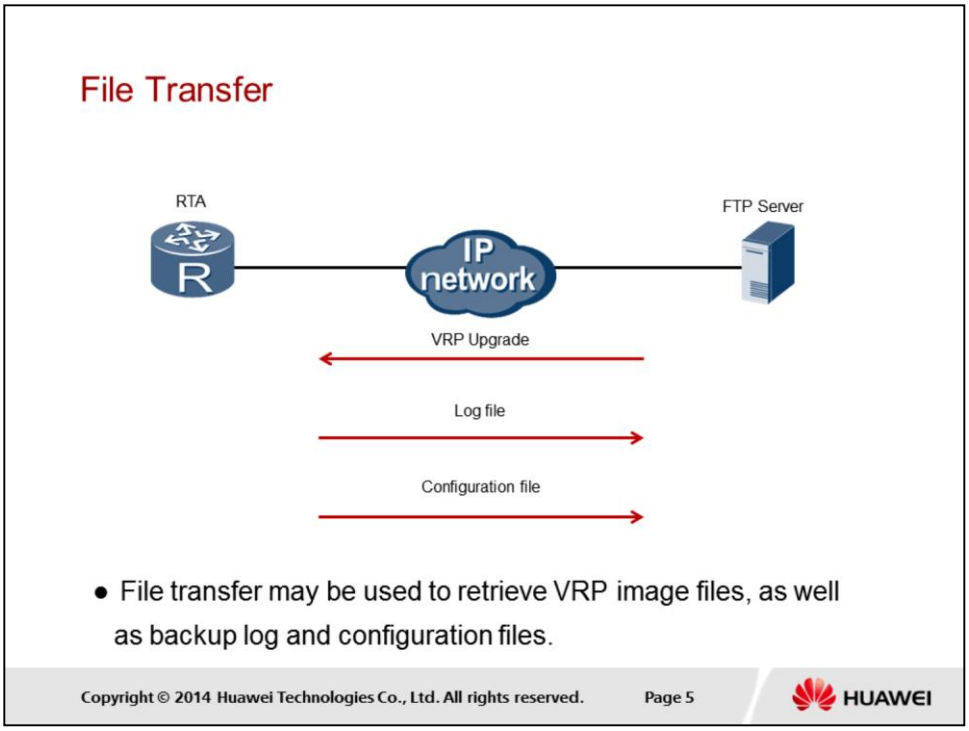
The VRP platform is constantly updated to maintain alignment with changes in technology and support new advancements to the hardware. The VRP image is generally defined by a VRP version and a product version number. Huawei ARG3 and Sx7 series products generally align with VRP version 5 to which different product versions are associated.

As the product version increases, so do the features that are supported by the version. The product version format includes a product code Vxxx , Rxxx denotes a major version release and Cxx a minor version release. If a service pack is used to patch the VRP product version, an SPC value may also be included in the VRP product version number. Typical examples of the VRP version upgrades for the AR2200E include:

Version 5.90 (AR2200 V200R001C00)

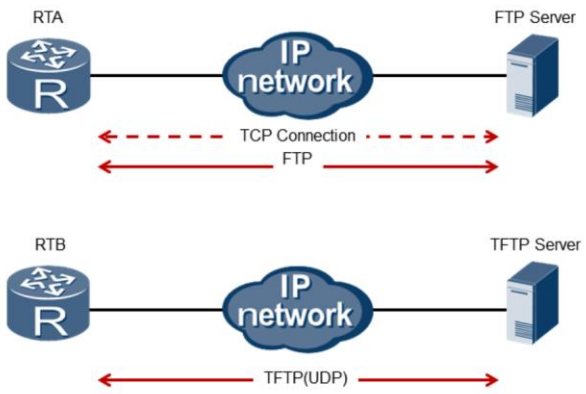
Version 5.110 (AR2200 V200R002C00)

Version 5.160 (AR2200 V200R007C00)



File transfer refers to the means by which files are sent to or retrieved from a remote server or storage location. Within the IP network this application can be implemented for a wide range of purposes. As part of effective practice, it is common for important files be duplicated and backed up within a remote storage location to prevent any loss that would affect critical systems operations. This includes files such as the VRP image of products which (should the existing image suffer loss through use of the format command or other forms of error), can be retrieved remotely and used to recover system operations. Similar principles apply for important configuration files and maintaining records of activity within devices stored in log files, which may be stored long term within the remote server.

File Transfer Methods



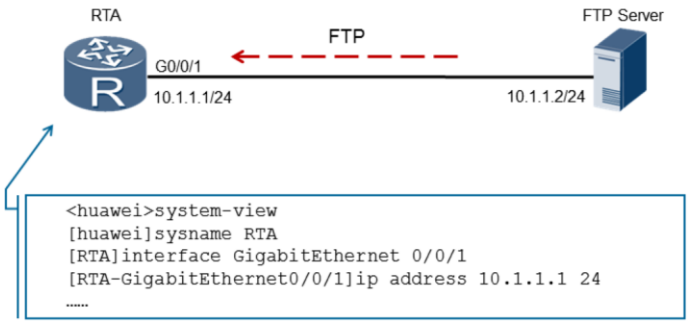
- Common forms of file transfer include FTP and TFTP, that respectively vary in the transport layer protocols used.

HUAWEI

FTP is a standard application protocol based on the TCP/IP protocol suite and used to transfer files between local clients and remote servers. FTP uses two TCP connections to copy a file from one system to another. The TCP connections are usually established in client-server mode, one for control (the server port number is 21) and the other for data transmission (the sever port number is 20). FTP as a file transfer protocol is used to control connections by issuing commands from the client (RTA) to the server and transmits replies from the server to the client, minimizing the transmission delay. In terms of data transmission, FTP transmits data between the client and server, maximizing the throughput.

Trivial File Transfer Protocol (TFTP) is a simple file transfer protocol over which a router can function as a TFTP client to access files on a TFTP server. Unlike FTP, TFTP has no complex interactive access interface and authentication control. Implementation of TFTP is based on the User Datagram Protocol (UDP). The client initiates the TFTP transfer. To download files, the client sends a read request packet to the TFTP server, receives packets from the server, and returns an acknowledgement to the server. To upload files, the client sends a write request packet to the TFTP server, sends packets to the server, and receives acknowledgement from the server.

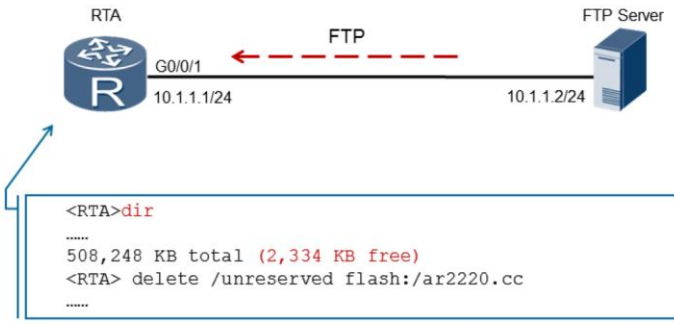
VRP Upgrade Process



HUAWEI

The example demonstrates how connection between an FTP server and client is established in order to retrieve a VRP image that can be used as part of the system upgrade process. Prior to any transfer of data, it is necessary to establish the underlying connectivity over which files can be transferred. This begins by providing suitable IP addressing for the client and the server. Where the devices are directly connected, interfaces can be applied that belong to the same network. Where devices belong to networks located over a large geographic area, devices must establish relevant IP addressing within their given networks and be able to discover a relevant network path over IP via which client/server connectivity can be established.

Storage Space Availability



- Where the storage capacity is inadequate for image transfer, older images and files can be removed.

HUAWEI

A user must determine for any system upgrade as to whether there is adequate storage space in which to store the file that is to be retrieved. The file system commands can be used to determine the current status of the file system, including which files are currently present within the file storage location of the device and also the amount of space currently available. Where the storage space is not adequate for file transfer, certain files can be deleted or uploaded to the FTP server in the event that they may still be required for future use.

The example demonstrates the use of the *delete* file system command to remove the existing image file. It should be noted that the system image, while deleted will not impact the current operation of the device as long as the device remains operational, therefore the device should not be powered off or restarted before a new VRP image file is restored within the storage location of the device, and set to be used during the next system startup.

Retrieving Files from an FTP Server


```
<RTA>ftp 10.1.1.2
Trying 10.1.1.2 ...
Press CTRL+K to abort
Connected to 10.1.1.2.
220 FTP service ready.
User(10.1.1.2:(none)):huawei
331 Password required for huawei.
Enter password:
230 User logged in.
[ftp]get vrp.cc
```


HUAWEI

The retrieving of files from an FTP server requires that a connection be established firstly before any file transfer can take place. Within the client device, the ftp service is initiated using the *ftp <ip address>* where the IP address relates to the address of the FTP server to which the client wishes to connect. FTP connections will be established using TCP, and requires authentication in the form of a username and password which is defined by the FTP server. Once authentication has been successfully achieved, the client will have established access to the FTP server and will be able to use a variety of commands to view existing files stored within the local current directory of the server.

Prior to file transmission, the user may be required to set the file type for which two formats exist, ASCII and Binary. ASCII mode is used for text, in which data is converted from the sender's character representation to "8-bit ASCII" before transmission, and then to the receiver's character representation. Binary mode on the other hand requires that the sender send each file byte for byte. This mode is often used to transfer image files and program files, and should be applied when sending or retrieving any VRP image file. In the example, the *get vrp.cc* command has been issued in order to retrieve the new VRP image located within the remote server.

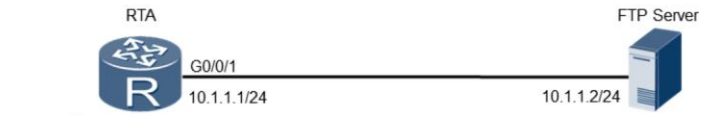
Retrieving Files from a TFTP Server


```
<RTA>tftp 10.1.1.2 get AR2220E-V200R007C00SPC600.cc
```

- A single command including the destination IP address is used to retrieve files from a TFTP server.

In the event that the client wishes to retrieve a VRP image from a TFTP server, a connection to the server need not first be established. Instead the client must define the path to the server within the command line, along with the operation that is to be performed. It should also be noted that the AR2200E & S5720 models serve as the TFTP client only and transfer files only in binary format. As can be seen from the example, the *get* command is applied for retrieval of the VRP image file from the TFTP server following the defining of the destination address of the TFTP server.

VRP Boot Management Process

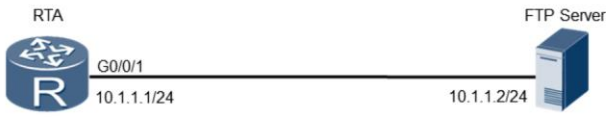



```
<RTA>startup system-software vrp.cc
Info: Succeeded in setting the software for booting system
<RTA>display startup
MainBoard:
  Configured startup system software:       flash:/ar2220.cc
  Startup system software:                  flash:/ar2220.cc
  Next startup system software:             vrp.cc
  Startup saved-configuration file:         NULL
  Next startup saved-configuration file:    NULL
  ......
```


The transfer of the VRP image file to the client once successfully achieved, requires that the image be enabled as the startup system software during the next system startup process. In order to change the system software version, the *startup system-software* command must be run and include the system software file to be used in the next startup. A system software file must use .cc as the file name extension, and the system software file used in the next startup cannot be that used in the current startup.

Additionally, the storage directory of a system software file must be the root directory, otherwise the file will fail to run. The *display startup* command should be used to verify that the change to the startup system software has been performed successfully. The output for the startup system software should show the existing VRP image, while the next startup system software should display the transferred VRP image that is now present within the root directory of the device.

Applying the Changes

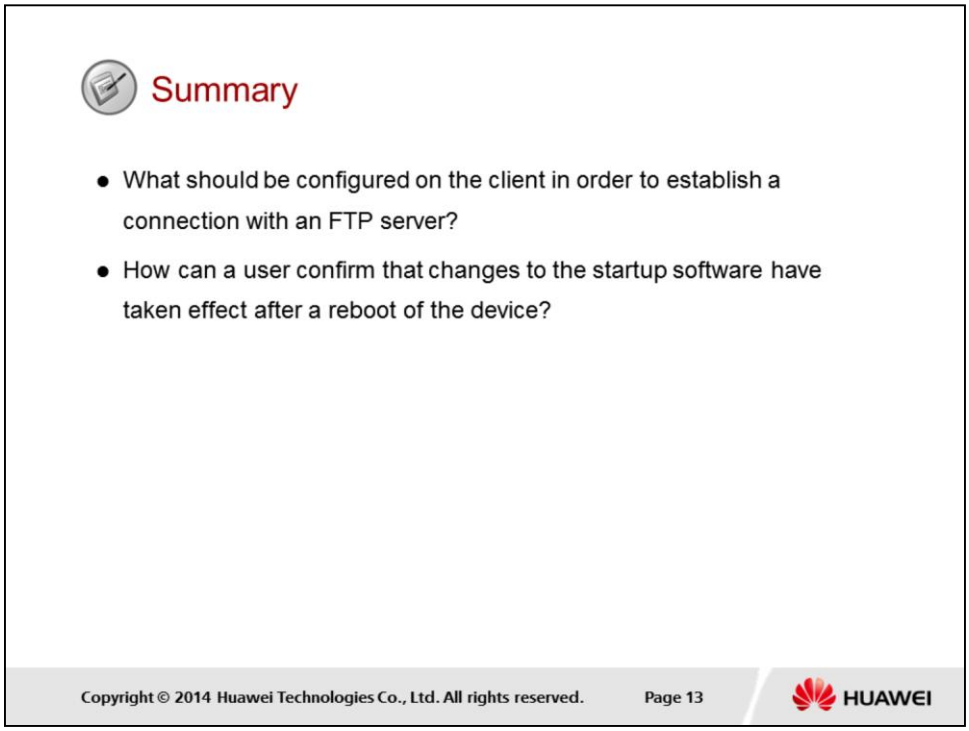



```
<RTA>reboot
Info: The system is now comparing the configuration, please
wait.
Warning: All the configuration will be saved to the
configuration file for the next startup, Continue?[Y/N]:n
System will reboot! Continue?[Y/N]:y
```

- The system must be restarted before the new image can take effect.

Confirmation of the startup system software allows for the safe initiation of the system software during the next system boot. In order to apply the changes and allow for the new system software to take effect, the device must be restarted. The *reboot* command can be used in order to initiate the system restart. During the reboot process, a prompt will be displayed requesting confirmation regarding whether the configuration file for the next system startup be saved.

In some cases, the saved-configuration file may be erased by the user in order to allow for a fresh configuration to be implemented. Should this have occurred, the user is expected define a response of ‘no’ at the ‘Continue?’ prompt. If the user chooses ‘yes’ at this point, the current-configuration will be rewritten to the saved-configuration file and applied once again during the next startup. If the user is unaware of the changes for which the save prompt is providing a warning, it is recommended that the user select ‘no’ or ‘n’ and perform a comparison of the saved and current configuration to verify the changes. For the reboot prompt, a response of ‘yes’ or ‘y’ is required to complete the reboot process.



1. A client device must have the capability to reach the FTP server over IP, requiring an IP address be configured on the interface via which the FTP server can be reached. This will allow a path to be validated to the FTP server at the network layer if one exists.
2. The user can run the configuration command *display startup* to validate that current startup system software (VRP) is active, identified by the .cc extension.

Thank you

www.huawei.com

Establishing a Single Switched Network

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

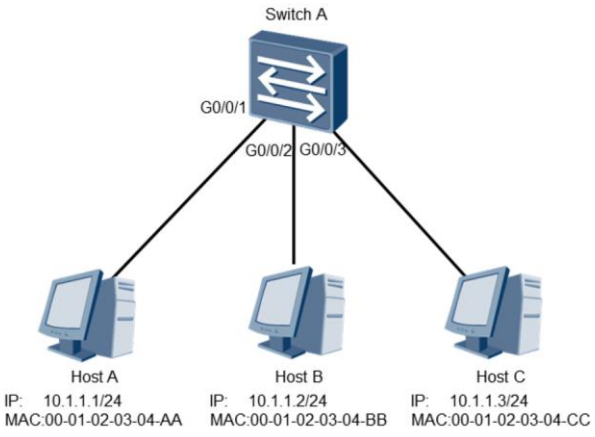
The introduction of a switching device as part of the enterprise network demonstrates how networks are able to expand beyond point-to-point connections, and shared networks in which collisions may occur. The behavior of the enterprise switch when introduced to the local area network is detailed along with an understanding of the handling of unicast and broadcast type frames, to demonstrate how switches enable networks to overcome the performance obstacles of shared networks.

Objectives

Upon completion of this section, trainees will be able to:

- Explain the decision making process of a link layer switch
- Configure parameters for negotiation on a link layer switch

Building a Single Switched Network

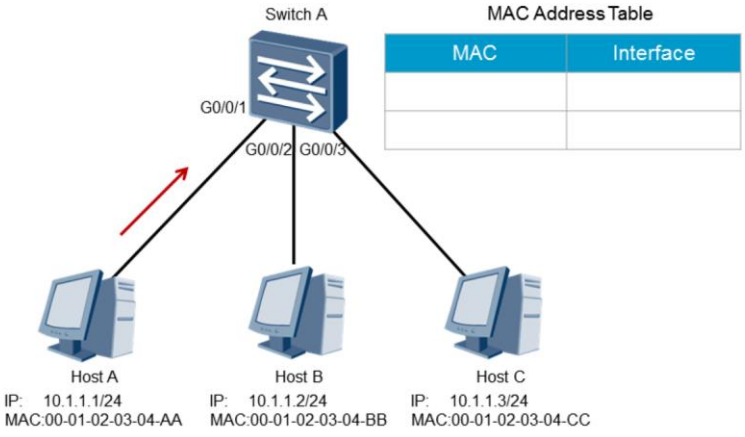


- Switches operate within the scope of the data link layer.

HUAWEI

As the enterprise network expands, multiple users need to be established as part of a multi-access network. The evolution of network technologies has seen a shift away from shared local networks, to networks which support multiple collision domains and support the use of 100BaseT forms of media that isolated the transmission and reception of data over separate wire pairs, thus eliminating the potential for collisions to occur and allowing higher full duplex transmission rates. The establishment of a switch brings the capability for increased port density to enable the connection of a greater number of end system devices within a single local area network. Each end system or host within a local area network is required to be connected as part of the same IP network in order for communication to be facilitated at the network layer. The IP address however is only relevant to the host systems since switch devices operate within the scope of the link layer and therefore rely on MAC addressing for frame forwarding.

The Initial State of The Switch

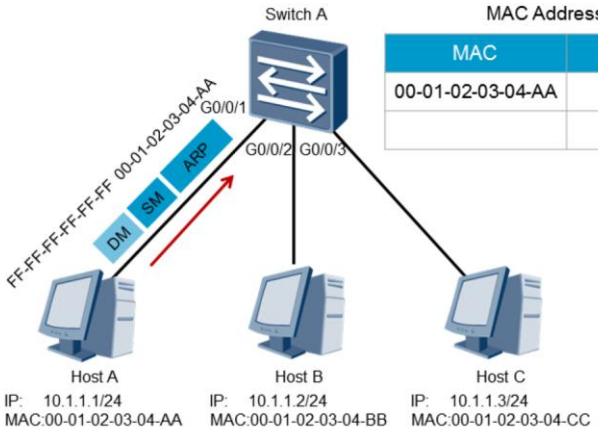


- Each switch uses a MAC table for making forwarding decisions.

As a link layer device, each switch relies on a MAC based table that provides association between a destination MAC address and the port interface via which a frame should be forwarded. This is commonly referred to as the MAC address table.

The initiation of a switch begins with the switch having no knowledge of end systems and how frames received from end systems should be forwarded. It is necessary that the switch build entries within the MAC address table to determine the path that each frame received should take in order to reach a given destination, so as to limit broadcast traffic within the local network. These path entries are populated in the MAC address table as a result of frames received from end systems. In the example, Host A has forwarded a frame to Switch A, which currently has no entries within its MAC address table.

MAC Address Learning

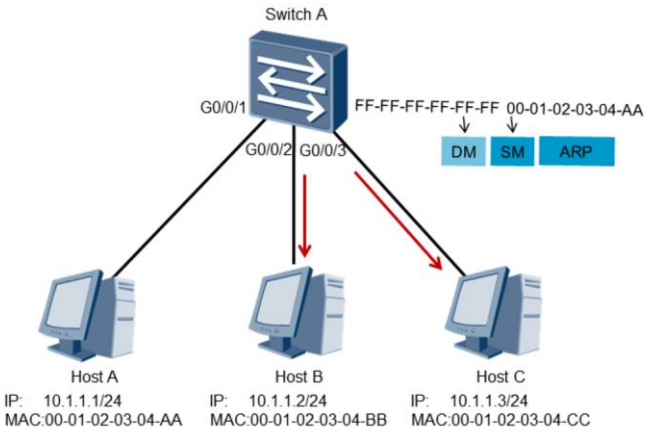


- The source MAC addresses of received frames are recorded.

HUAWEI

The frame that is forwarded from Host A contains a broadcast MAC address entry in the destination address field of the frame header. The source address field contains the MAC address of the peering device, in this case Host A. This source MAC address is used by the switch in order to populate the MAC address table, by associating the MAC entry in the source address field with the switch port interface upon which the frame was received. The example demonstrates how the MAC address is associated with the port interface to allow any returning traffic to this MAC destination to be forwarded directly via the associated interface.

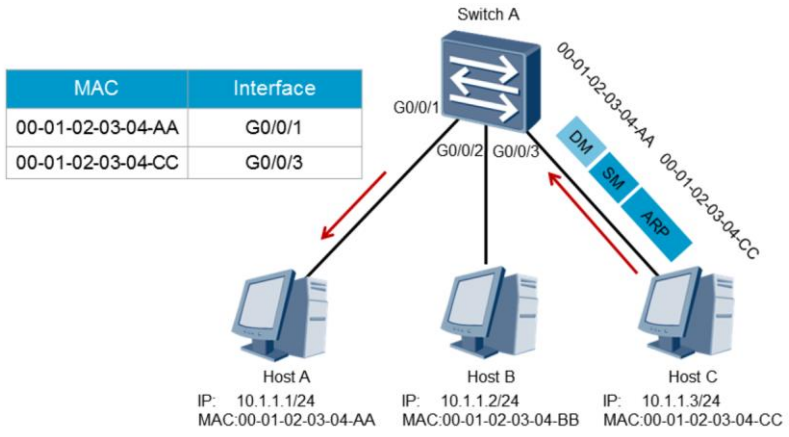
Forwarding The First Data



- Frames destined for unknown link layer destinations are flooded.

The general behavior of an ARP request involves the frame being flooded to all intended destinations primarily due to the MAC broadcast (FF:FF:FF:FF:FF:FF) that represents the current destination. The switch is therefore responsible for forwarding this frame out of every port interface with exception to the port interface on which the frame was received, in an attempt to locate the intended IP destination as listed within the ARP header for which an ARP reply can be generated. As demonstrated in the example, individual frames are flooded from the switch via port interfaces G0/0/2 and G0/0/3 towards hosts B and host C respectively.

The Destination Reply



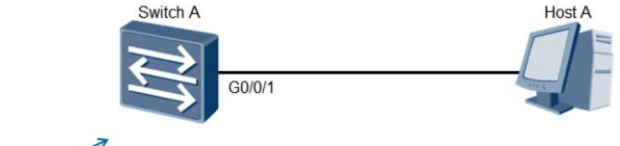
- Frames are forwarded to destinations based on the MAC table.

HUAWEI

As a result of the ARP request header, the receiving host is able to determine that the ARP header is intended for the IP destination of 10.1.1.3, along with the local source address (MAC) from which the frame originated, and use this information to generate a unicast reply. The information regarding Host A is associated with the IP address of 10.1.1.3 and stored within the MAC address table of Host C. In doing so, the generation of broadcast traffic is minimized, thereby reducing the number of interrupts to local destinations as well as reduction of the number of frames propagating the local network.

Once the frame is received from Host C by Switch A, the switch will populate the MAC address table with the source MAC address of the frame received, and associate it with the port interface on which the frame was received. The switch then uses the MAC address table to perform a lookup, in order to discover the forwarding interface, based on the destination MAC address of the frame. In this case the MAC address of the frame refers to Host A, for which an entry now exists via interface G0/0/1, allowing the frame to be forwarded to the known destination.

Basic Configuration



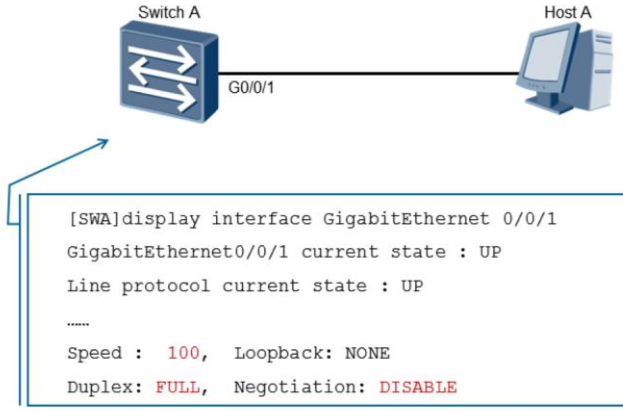

```
<SWA>system-view
Enter system view, return user view with Ctrl+Z.
[SWA]interface GigabitEthernet 0/0/1
[SWA-GigabitEthernet0/0/1]undo negotiation auto
[SWA-GigabitEthernet0/0/1]duplex full
[SWA-GigabitEthernet0/0/1]speed 100
```


Early Ethernet systems operated based on a 10Mbps half duplex mode and applied mechanisms such as CSMA/CD to ensure system stability. The transition to a twisted pair medium gave rise to the emergence of full-duplex Ethernet, which greatly improved Ethernet performance and meant two forms of duplex could be negotiated. The auto-negotiation technology allows newer Ethernet systems to be compatible with earlier Ethernet systems.

In auto-negotiation mode, interfaces on both ends of a link negotiate their operating parameters, including the duplex mode, rate, and flow control. If the negotiation succeeds, the two interfaces work with the same operating parameters. In some cases however it is necessary to manually define the negotiation parameters, such as where Gigabit Ethernet interfaces that are working in auto-negotiation mode are connected via a 100 Mbps network cable. In such cases, negotiation between the interfaces will fail.

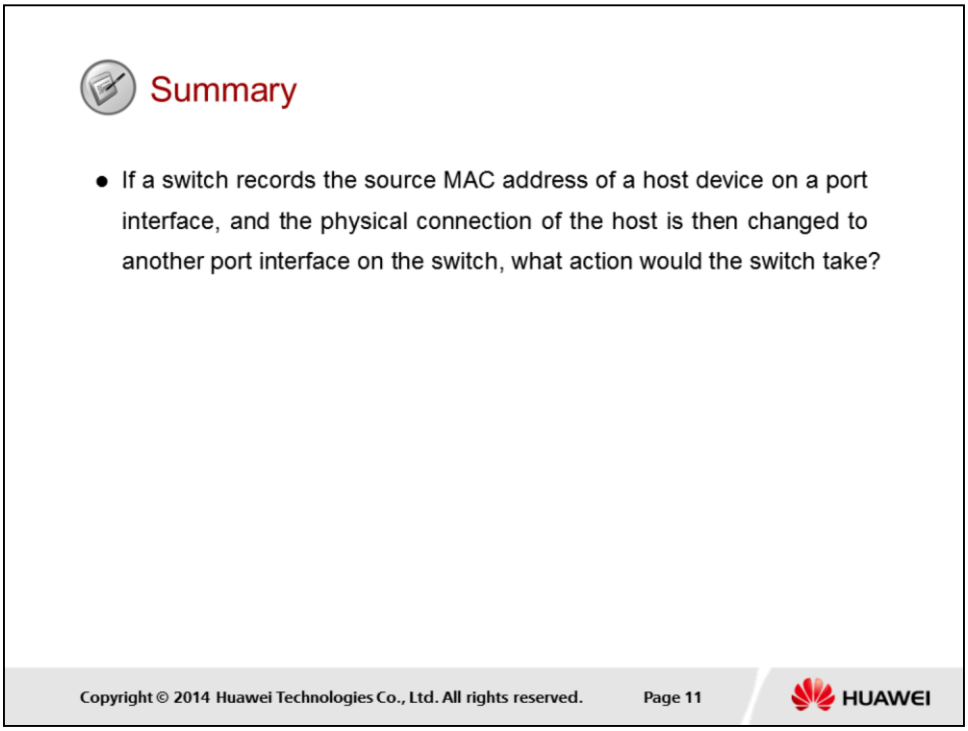
Due to different product models, HUAWEI switches may not support the change port duplex mode, see the product manual.

Basic Configuration Verification



HUAWEI

In the event that the configuration parameters for negotiation are changed from using auto negotiation, the defined parameters should be checked using the *display interface <interface>* command to verify that the negotiated parameters allow for the link layer interface negotiation to be successful. This is verified by the line protocol current state being displayed as UP. The displayed information reflects the current parameter settings for an interface.



1. When a host or other end system is connected to a switch port interface, a gratuitous ARP is generated that is designed to ensure that IP addresses remain unique within a network segment. The gratuitous ARP message however also provides the switch with information regarding the MAC address of the host, which is then included in the MAC address table and associated with the port interface on which the host is connected.

   If the physical connection of a host connected to a switch port interface is removed, the switch will discover the physical link is down and remove the MAC entry from the MAC address table. Once the medium is connected to another port interface, the port will detect that the physical link is active and a gratuitous ARP will be generated by the host, allowing the switch to discover and populate the MAC address table with the MAC address of the connected host.

Thank you

www.huawei.com

Spanning Tree Protocol

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

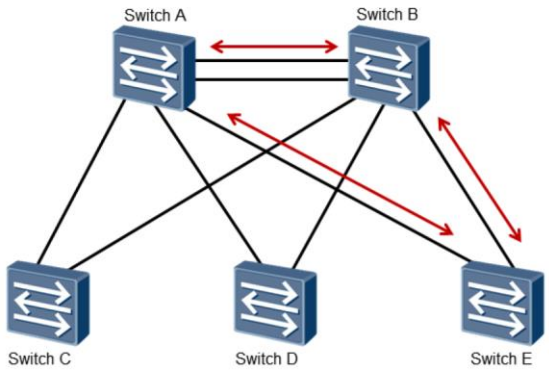
As the enterprise network expands, multi-switched networks are introduced to provide link layer communication between a growing number of end systems. As new interconnections are formed between multiple enterprise switches, new opportunities for building ever resilient networks are made possible, however the potential for switching failure as a result of loops becomes ever more likely. It is necessary that the spanning tree protocol (STP) therefore be understood in terms of behavior in preventing switching loops, and how it can be manipulated to suit enterprise network design and performance.

Objectives

Upon completion of this section, trainees will be able to:

- Describe the issues faced when using a multi-switched network.
- Explain the loop prevention process of the spanning tree protocol.
- Configure parameters for managing the STP network design.

Layer 2 Redundancy



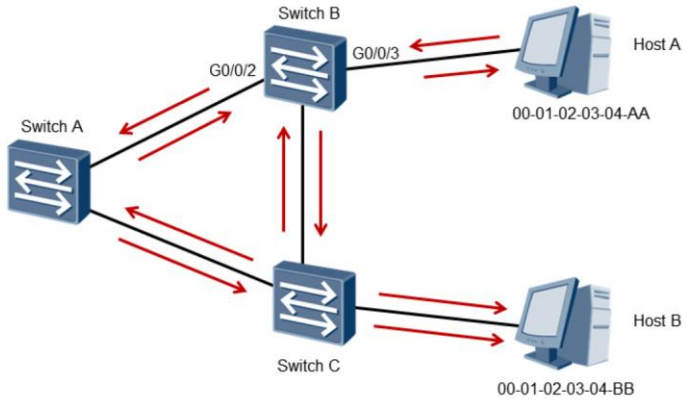
- Redundancy in a switching network minimizes connection failure but generates potential switching loops.

HUAWEI

Enterprise growth results in the commissioning of multiple switches in order to support the interconnectivity of end systems and services required for daily operations. The interconnection of multiple switches however brings additional challenges that need to be addressed. Switches may be established as single point-to-point links via which end systems are able to forward frames to destinations located via other switches within the broadcast domain. The failure however of any point-to-point switch link results in the immediate isolation of the downstream switch and all end systems to which the link is connected. In order to resolve this issue, redundancy is highly recommended within any switching network.

Redundant links are therefore generally used on an Ethernet switching network to provide link backup and enhance network reliability. The use of redundant links, however, may produce loops that cause the communication quality to drastically deteriorate, and major interruptions to the communication service to occur.

Broadcast Storms



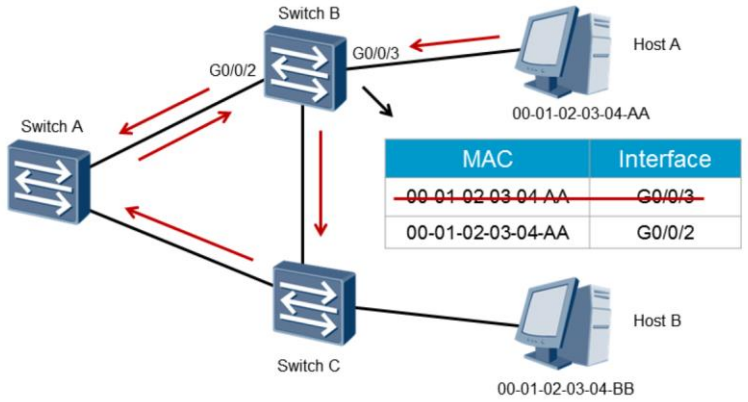
- Switching loops allow for broadcast storms to occur and duplication of frames to be received by end stations.

HUAWEI

One of the initial effects of redundant switching loops comes in the form of broadcast storms. This occurs when an end system attempts to discover a destination for which neither itself nor the switches along the switching path are aware of. A broadcast is therefore generated by the end system which is flooded by the receiving switch.

The flooding effect means that the frame is forwarded via all interfaces with exception to the interface on which the frame was received. In the example, Host A generates a frame, which is received by Switch B which is subsequently forwarded out of all other interfaces. An instance of the frame is received by the connected switches A and C, which in turn flood the frame out of all other interfaces. The continued flooding effect results in both Switch A and Switch C flooding instances of the frame from one switch to the other, which in turn is flooded back to Switch B, and thus the cycle continues. In addition, the repeated flooding effect results in multiple instances of the frame being received by end stations, effectively causing interrupts and extreme switch performance degradation.

MAC Instability

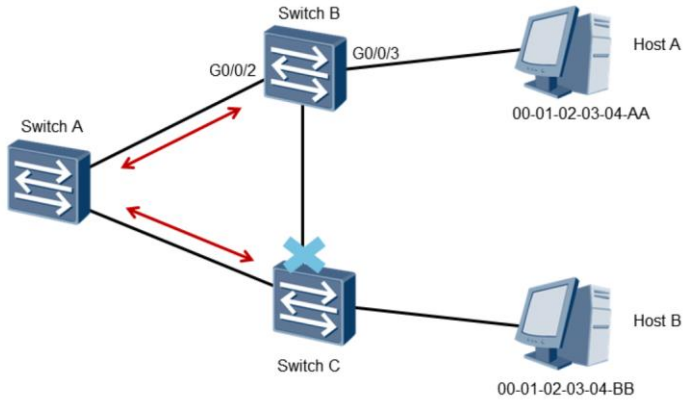


- Receiving previously forwarded frames generates false MAC entries, and instability within the MAC address table.

Switches must maintain records of the path via which a destination is reachable. This is identified through association of the source MAC address of a frame with the interface on which the frame was received. Only one instance of a MAC address can be stored within the MAC address table of a switch, and where a second instance of the MAC address is received, the more recent information takes precedence.

In the example, Switch B updates the MAC address table with the MAC address of Host A and associates this source with interface G0/0/3, the port interface on which the frame was received. As frames are uncontrollably flooded within the switching network, a frame is again received with the same source MAC address as Host A, however this time the frame is received on interface G0/0/2. Switch B must therefore assume that the host that was originally reachable via interface G0/0/3 is now reachable via G0/0/2, and will update the MAC address table accordingly. The result of this process leads to MAC instability and continues to occur endlessly between both the switch port interfaces connecting to Switch A and Switch C since frames are flooded in both directions as part of the broadcast storm effect.

Resolving Layer 2 Redundancy Issues

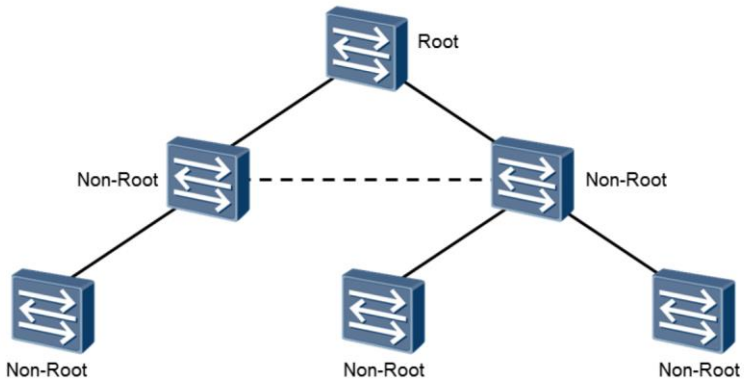


- Loops are eliminated by restricting traffic flow over redundant paths.

HUAWEI

The challenge for the switching network lies in the ability to maintain switching redundancy to avoid isolation of end systems in the event of switch system or link failure, and the capability to avoid the damaging effects of switching loops within a switching topology which implements redundancy. The resulting solution for many years has been to implement the spanning tree protocol (STP) in order to prevent the effects of switching loops. Spanning tree works on the principle that redundant links be logically disabled to provide a loop free topology, whilst being able to dynamically enable secondary links in the event that a failure along the primary switching path occurs, thereby fulfilling the requirement for network redundancy within a loop free topology. The switching devices running STP discover loops on the network by exchanging information with one another, and block certain interfaces to cut off loops. STP has continued to be an important protocol for the LAN for over 20 years.

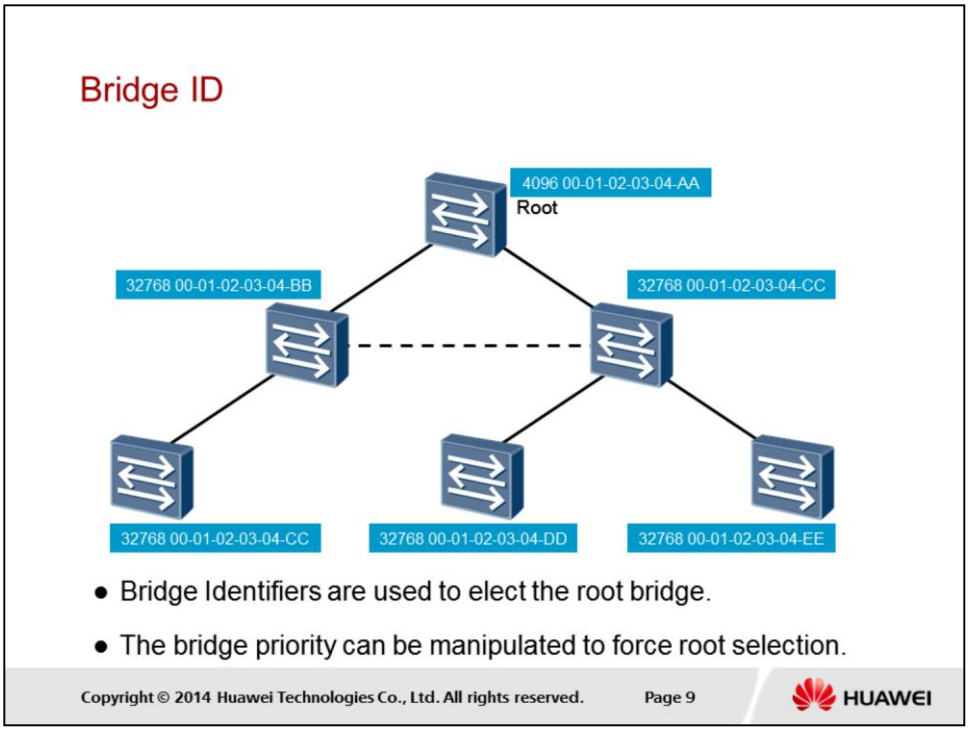
The Spanning Tree Root Bridge



- An inverted tree architecture is created as a result of STP.
- The root bridge represents the base of the spanning tree.

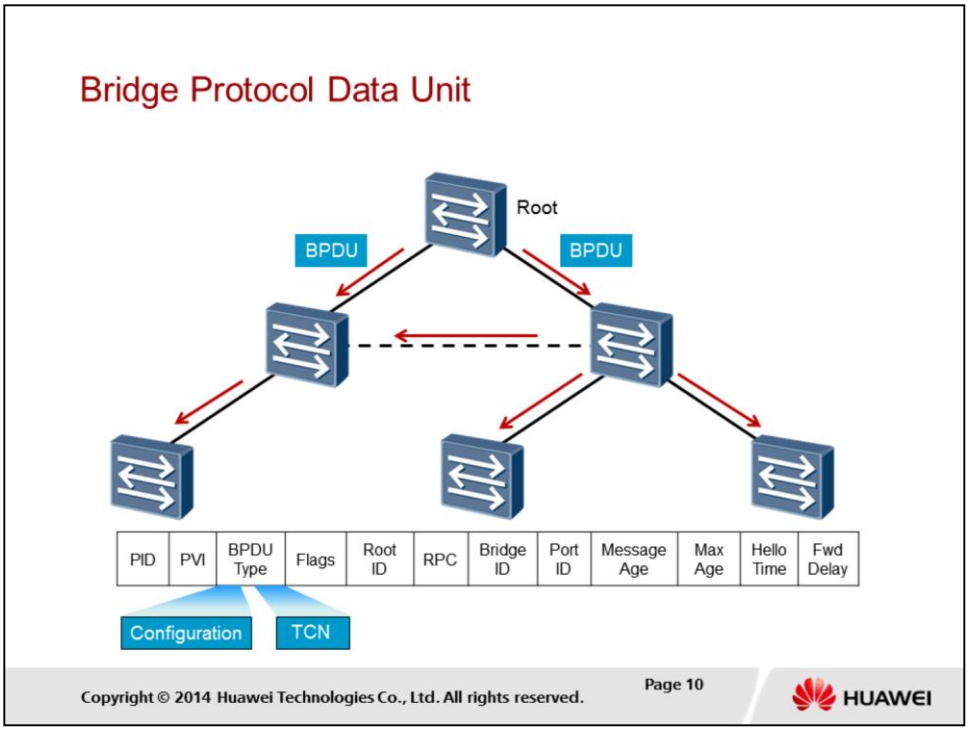
HUAWEI

The removal of any potential for loops serves as the primary goal of spanning tree for which an inverted tree type architecture is formed. At the base of this logical tree is the root bridge/switch. The root bridge represents the logical center but not necessarily the physical centre of the STP-capable network. The designated root bridge is capable of changing dynamically with the network topology, as in the event where the existing root bridge fails to continue to operate as the root bridge. Non-root bridges are considered to be downstream from the root bridge and communication to non-root bridges flows from the root bridge towards all non-root bridges. Only a single root bridge can exist in a converged STP-capable network at any one time.



Discovery of the root bridge for an STP network is a primary task performed in order to form the spanning tree. The STP protocol operates on the basis of election, through which the role of all switches is determined. A bridge ID is defined as the means by which the root bridge is discovered. This comprises of two parts, the first being a 16 bit bridge priority and the second, a 48 bit MAC address.

The device that is said to contain the highest priority (smallest bridge ID) is elected as the root bridge for the network. The bridge ID comparison takes into account initially the bridge priority, and where this priority value is unable to uniquely identify a root bridge, the MAC address is used as a tie breaker. The bridge ID can be manipulated through alteration to the bridge priority as a means of enabling a given switch to be elected as the root bridge, often in support of an optimized network design.

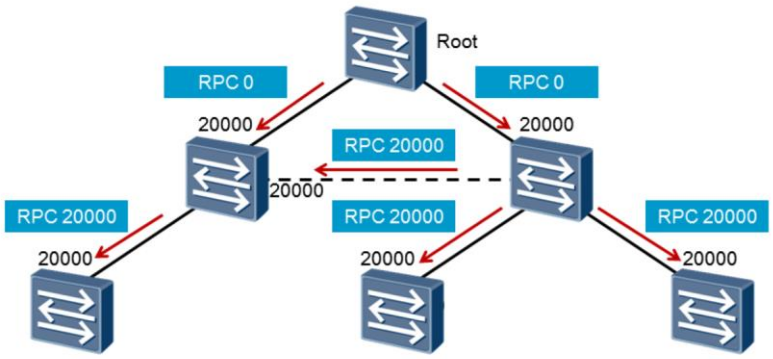


The spanning tree topology relies on the communication of specific information to determine the role and status of each switch in the network. A Bridge Protocol Data Unit (BPDU) facilitates communication within a spanning tree network. Two forms of BPDU are used within STP. A Configuration BPDU is initially created by the root and propagated downstream to ensure all non-root bridges remain aware of the status of the spanning tree topology and importantly, the root bridge. The TCN BPDU is a second form of BPDU, which propagates information in the upstream direction towards the root and shall be introduced in more detail as part of the topology change process.

Bridge Protocol Data Units are not directly forwarded by switches, instead the information that is carried within a BPDU is often used to generate a switches own BPDU for transmission. A Configuration BPDU carries a number of parameters that are used by a bridge to determine primarily the presence of a root bridge and ensure that the root bridge remains the bridge with the highest priority. Each LAN segment is considered to have a designated switch that is responsible for the propagation of BPDU downstream to non-designated switches.

The Bridge ID field is used to determine the current designated switch from which BPDU are expected to be received. The BPDU is generated and forwarded by the root bridge based on a Hello timer, which is set to 2 seconds by default. As BPDU are received by downstream switches, a new BPDU is generated with locally defined parameters and forwarded to all non-designated switches for the LAN segment.

Path Cost



- Root path cost is carried in the BPDU and used to determine the shortest path to the root.

HUAWEI

Another feature of the BPDU is the propagation of two parameters relating to path cost. The root path cost (RPC) is used to measure the cost of the path to the root bridge in order to determine the spanning tree shortest path, and thereby generate a loop free topology. When the bridge is the root bridge, the root path cost is 0.

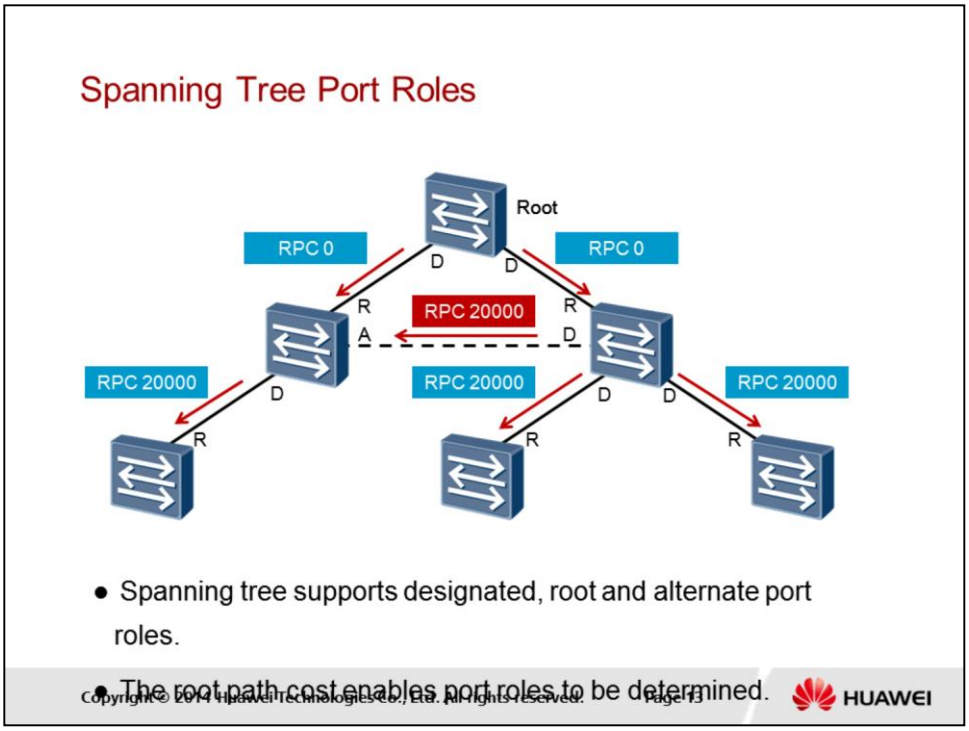
The path cost (PC) is a value associated with the root port, which is the port on a downstream switch that connects to the LAN segment, on which a designated switch or root bridge resides. This value is used to generate the root path cost for the switch, by adding the path cost to the RPC value that is received from the designated switch in a LAN segment, to define a new root path cost value. This new root path cost value is carried in the BPDU of the designated switch and is used to represent the path cost to the root.

Path Cost Standards

Port Speed	802.1D	802.1t	Path Cost Legacy
10 Mbps	99	1999999	1999
100 Mbps	18	199999	199
1 Gbps	4	20000	20
10 Gbps	2	2000	2

- STP supports various path cost standards
- The 802.1t is the default standard used by Huawei switches

Huawei Sx7 series switches support a number of alternative path cost standards that can be implemented based on enterprise requirements, such as where a multi vendor switching network may exist. The Huawei Sx7 series of switches use the 802.1t path cost standard by default, providing a stronger metric accuracy for path cost calculation.

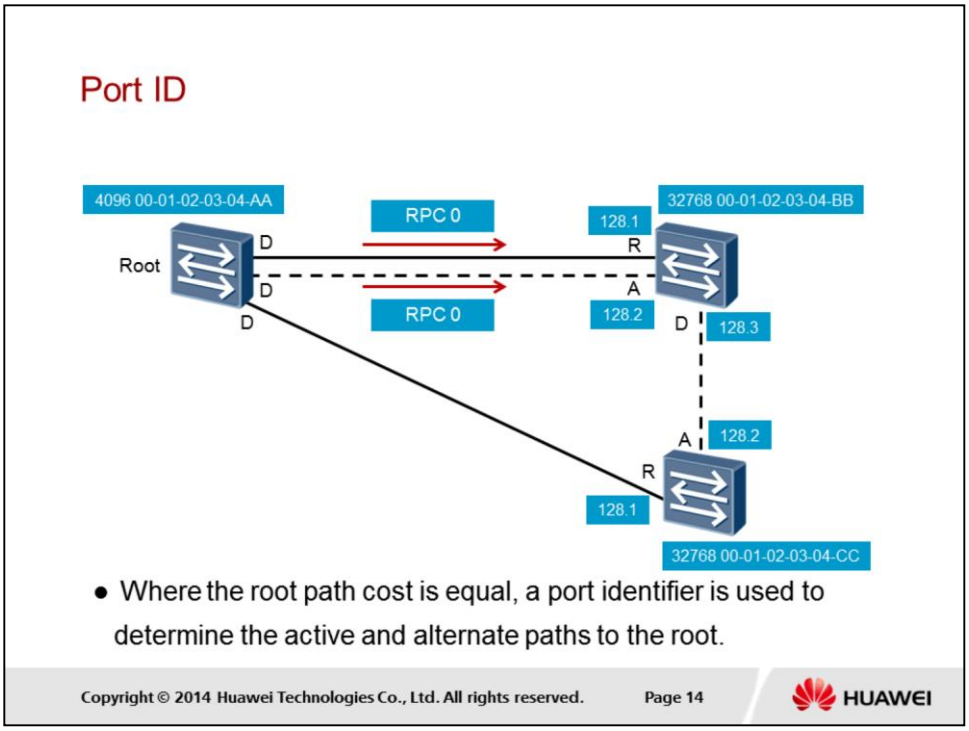


A converged spanning tree network defines that each interface be assigned a specific port role. Port roles are used to define the behavior of port interfaces that participate within an active spanning tree topology. For the spanning tree protocol, three port roles of designated, root and alternate are defined.

The designated port is associated with a root bridge or a designated bridge of a LAN segment and defines the downstream path via which Configuration BPDU are forwarded. The root bridge is responsible for the generation of configuration BPDU to all downstream switches, and thus root bridge port interfaces always adopt the designated port role.

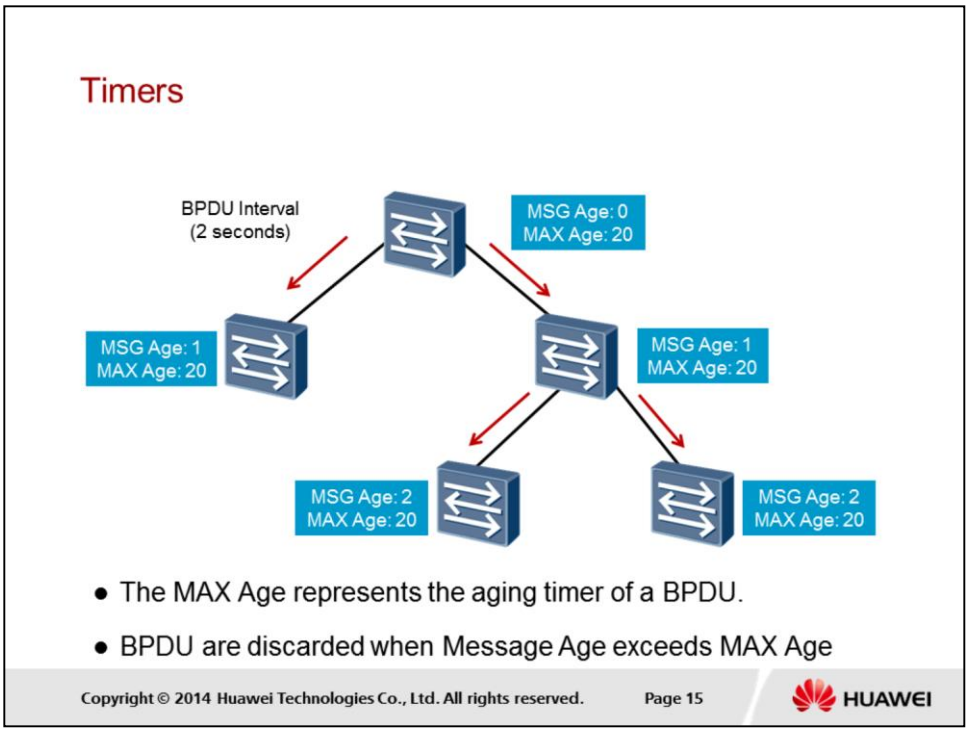
The root port identifies the port that offers the lowest cost path to the root, based on the root path cost. The example demonstrates the case where two possible paths exist back to the root, however only the port that offers the lowest root path cost is assigned as the root port. Where two or more ports offer equal root path costs, the decision of which port interface will be the root port is determined by comparing the bridge ID in the configuration BPDU that is received on each port.

Any port that is not assigned a designated or root port role is considered an alternate port, and is able to receive BPDU from the designated switch for the LAN segment for the purpose of monitoring the status of the redundant link, but will not process the received BPDU. The IEEE 802.1D-1990 standard for STP originally defined this port role as backup, however this was amended to become the alternate port role within the IEEE 802.1D-1998 standards revision.



The port ID represents a final means for determining port roles alongside the bridge ID and root path cost mechanism. In scenarios where two or more ports offer a root path cost back to the root that is equal and for which the upstream switch is considered to have a bridge ID that is equal, primarily due to the upstream switch being the same switch for both paths, the port ID must be applied to determine the port roles.

The port ID is tied to each port and comprises of a port priority and a port number that associates with the port interface. The port priority is a value in the range of 0 to 240, assigned in increments of 16, and represented by a value of 128 by default. Where both port interfaces offer an equal port priority value, the unique port number is used to determine the port roles. The highest port identifier (the lowest port number) represents the port assigned as the root port, with the remaining port defaulting to an alternate port role.

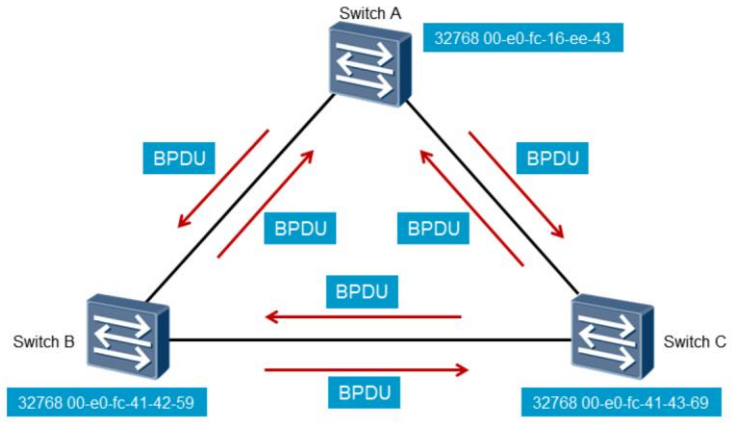


The root bridge is responsible for the generation of configuration BPDU based on a BPDU interval that is defined by a Hello timer. This Hello timer by default represents a period of 2 seconds. A converged spanning tree network must ensure that in the event of a failure within the network, that switches within the STP enabled network are made aware of the failure. A Max Age timer is associated with each BDPU and represents life span of a BPDU from the point of conception by the root bridge, and ultimately controls the validity period of a BDPU before it is considered obsolete. This MAX Age timer by default represents a period of 20 seconds.

Once a configuration BPDU is received from the root bridge, the downstream switch is considered to take approximately 1 second to generate a new BPDU, and propagate the generated BPDU downstream. In order to compensate for this time, a message age (MSG Age) value is applied to each BPDU to represent the offset between the MAX Age and the propagation delay, and for each switch this message age value is incremented by 1.

As BPDU are propagated from the root bridge to the downstream switches the MAX Age timer is refreshed. The MAX Age timer counts down and expires when the MAX Age value exceeds the value of the message age, to ensure that the lifetime of a BPDU is limited to the MAX Age, as defined by the root bridge. In the event that a BPDU is not received before the MAX Age timer expires, the switch will consider the BPDU information currently held as obsolete and assume an STP network failure has occurred.

Root Election Process



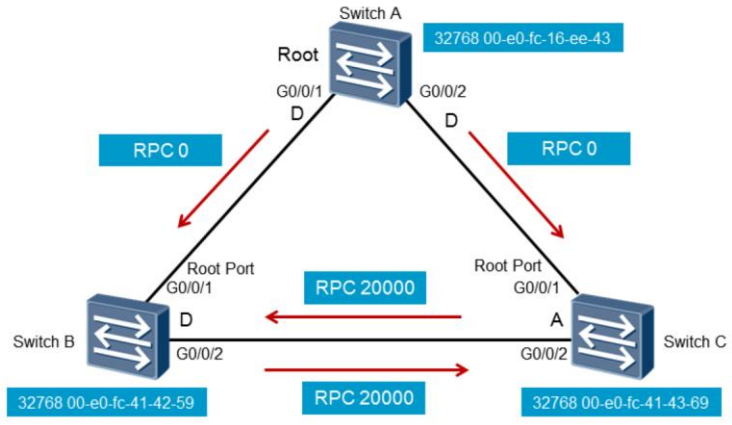
- All STP switches advertise BDPU to peers with self as root.

HUAWEI

The spanning tree convergence process is an automated procedure that initiates at the point of switch startup. All switches at startup assume the role of root bridge within the switching network. The default behavior of a root bridge is to assign a designated port role to all port interfaces to enable the forwarding of BPDU via all connected port interfaces. As BPDU are received by peering switches, the bridge ID will be compared to determine whether a better candidate for the role of root bridge exists. In the event that the received BPDU contains an inferior bridge ID with respect to the root ID, the receiving switch will continue to advertise its own configuration BPDU to the neighboring switch.

Where the BDPU is superior, the switch will acknowledge the presence of a better candidate for the role of root bridge, by ceasing to propagate BPDU in the direction from which the superior BPDU was received. The switch will also amend the root ID field of it’s BPDU to advertise the bridge ID of the root bridge candidate as the current new root bridge.

Port Role Establishment Process

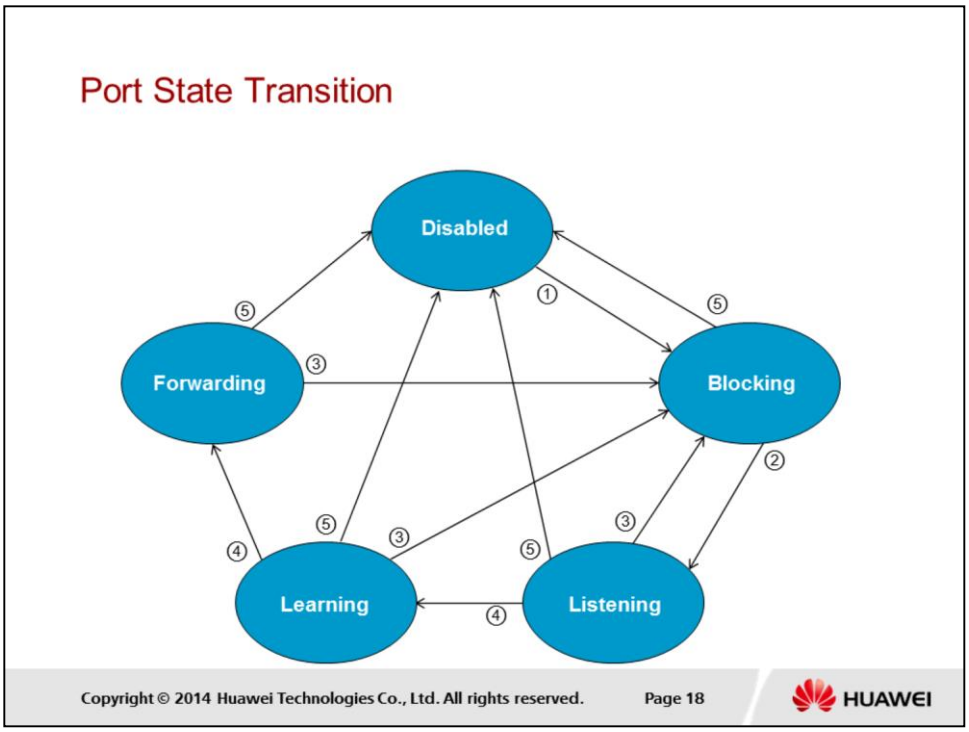


- The Bridge ID and Root Path Cost are used to elect port roles.

HUAWEI

An elected root bridge, once established will generate configuration BPDU to all other non-root switches. The BPDU will carry a root path cost that will inform downstream switches of the cost to the root, to allow for the shortest path to be determined. The root path cost carried in the BPDU that is generated by the root bridge always has a value of 0. The receiving downstream switches will then add this cost to the path cost of the port interfaces on which the BPDU was received, and from which a switch is able to identify the root port.

In the case where equal root path costs exist on two or more LAN segments to the same upstream switch, the port ID is used to discover the port roles. Where an equal root path cost exists between two switches as in the given example, the bridge ID is used to determine which switch represents the designated switch for the LAN segment. Where the switch port is neither a root port nor designated port, the port role is assigned as alternate.

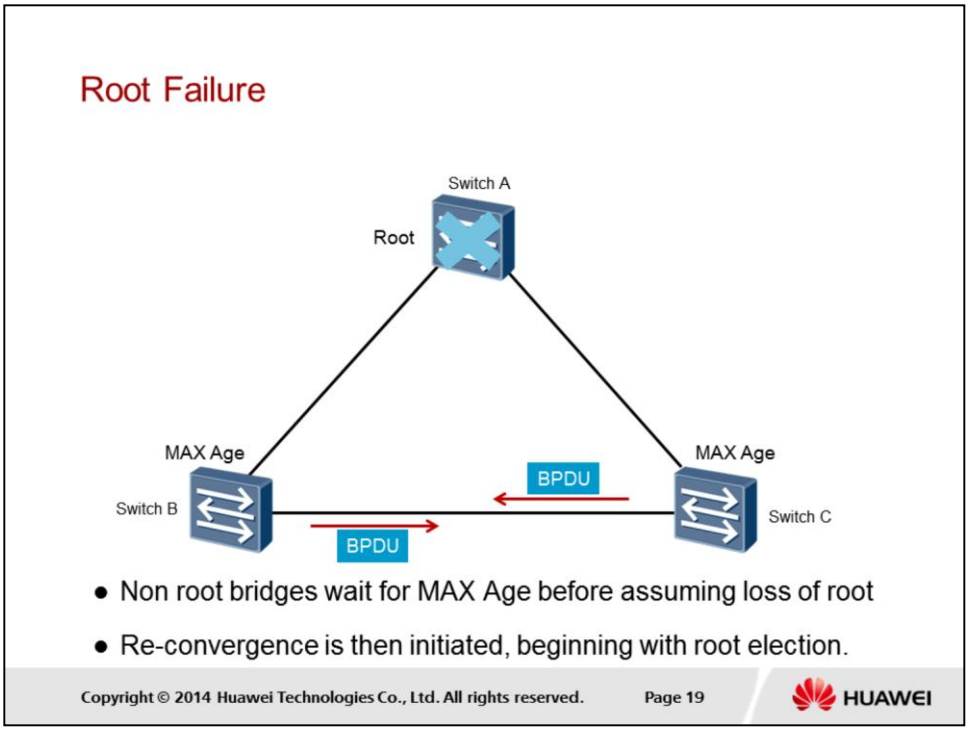


As part of the root bridge and port role establishment, each switch will progress through a number of port state transitions. Any port that is administratively disabled will be considered to be in the disabled state. Enabling of a port in the disabled state will see a state transition to the blocking state ①.

Any port considered to be in a blocking state is unable to forward any user traffic, but is capable of receiving BPDU frames. Any BPDU received on a port interface in the blocking state will not be used to populate the MAC address table of the switch, but instead to determine whether a transition to the listening state is necessary. The listening state enables communication of BPDU information, following negotiation of the port role in STP ②, but maintains restriction on the populating of the MAC address table with neighbor information.

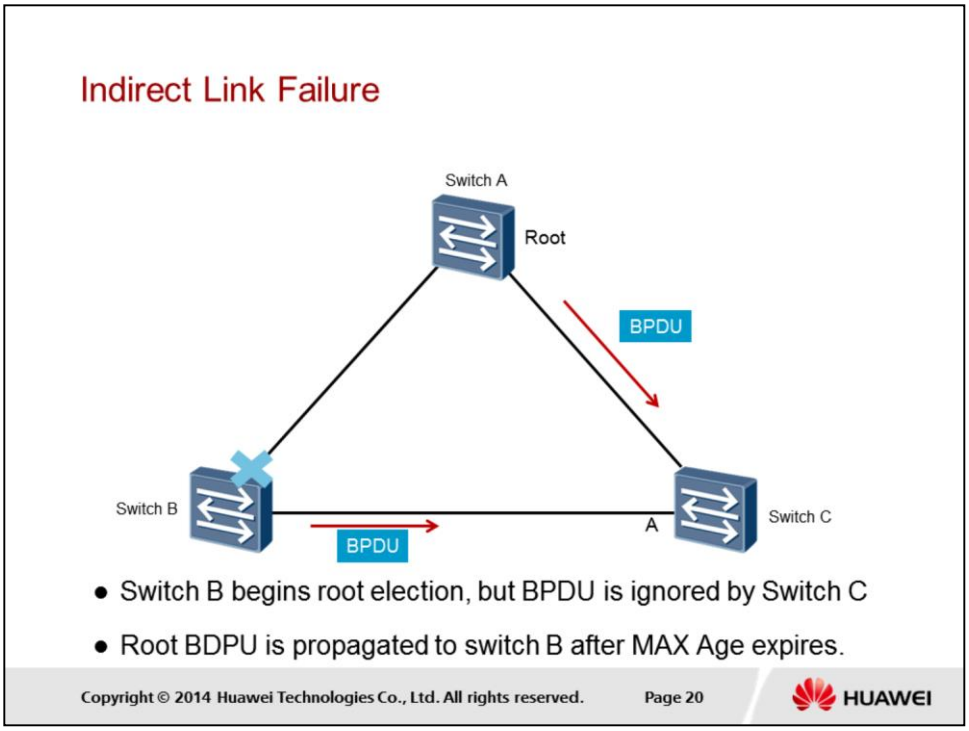
A transition to the blocking state from the listening or other states ③ may occur in the event that the port is changed to an alternate port role. The transition between listening to learning and learning to forwarding states ④ is greatly dependant on the forward delay timer, which exists to ensure that any propagation of BDPU information to all switches in the spanning tree topology is achievable before the state transition occurs.

The learning state maintains the restriction of user traffic forwarding to ensure prevention of any switching loops however allows for the population of the MAC address table throughout the spanning tree topology to ensure a stable switching network. Following a forward delay period, the forwarding state is reached. The disabled state is applicable at any time during the state transition period through manual intervention (i.e. the *shutdown* command) ⑤.



Events that cause a change in the established spanning tree topology may occur in a variety of ways, for which the spanning tree protocol must react to quickly re-establish a stable and loop free topology. The failure of the root bridge is a primary example of where re-convergence is necessary. Non-root switches rely on the intermittent pulse of BPDU from the root bridge to maintain their individual roles as non-root switches in the STP topology. In the event that the root bridge fails, the downstream switches will fail to receive a BPDU from the root bridge and as such will also cease to propagate any BPDU downstream. The MAX Age timer is typically reset to the set value (20 seconds by default) following the receipt of each BPDU downstream.

With the loss of any BPDU however, the MAX Age timer begins to count down the lifetime for the current BPDU information of each non-root switch, based on the (MAX Age – MSG Age) formula. At the point at which the MSG Age value is greater than the MAX Age timer value, the BPDU information received from the root becomes invalid, and the non-root switches begin to assume the role of root bridge. Configuration BPDU are again forwarded out of all active interfaces in a bid to discover a new root bridge. The failure of the root bridge invokes a recovery duration of approximately 50 seconds due to the Max Age + 2x Forward Delay convergence period.

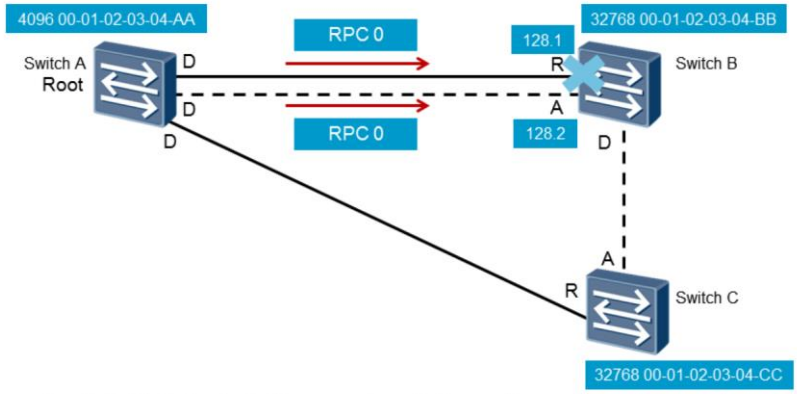


In the case of an indirect link failure, a switch loses connection with the root bridge due to a failure of the port or media, or due possibly to manual disabling of the interface acting as the root port. The switch itself will become immediately aware of the failure, and since it only receives BPDU from the root in one direction, will assume immediate loss of the root bridge, and assert its position as the new root bridge.

From the example, switch B begins to forward BPDU to switch C to notify of the position of switch B as the new root bridge, however switch C continues to receive BPDU from the original root bridge and therefore ignores any BPDU from switch B. The alternate port will begin to age its state through the MAX Age timer, since the interface no longer receives BPDU containing the root ID of the root bridge.

Following the expiry of the MAX Age timer, switch C will change the port role of the alternate port to that of a designated port and proceed to forward BPDU from the root towards switch B, which will cause the switch to concede its assertion as the root bridge and converge its port interface to the role of root port. This represents a partial topology failure however due to the need to wait for a period equivalent to MAX Age + 2x forward delay, full recovery of the STP topology requires approximately 50 seconds.

Direct Link Failure



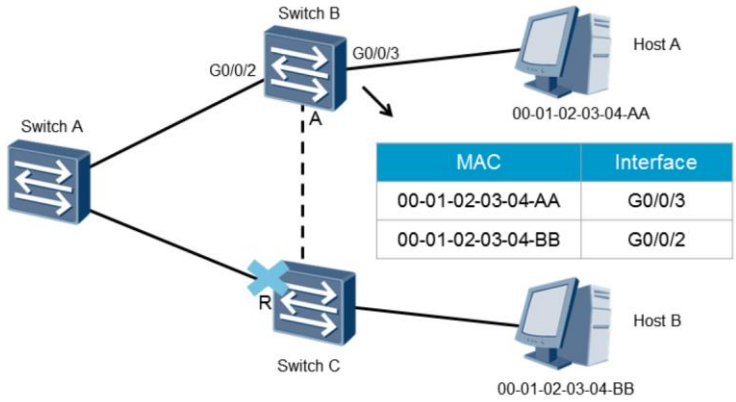
- Switch B detects failure and switches alternate port to root port.
- STP converges after 2x forward delay (30 seconds by default).

HUAWEI

A final scenario involving spanning tree convergence recovery occurs where multiple LAN segments are connected between two switch devices for which one is currently the active link while the other provides an alternate path to the root. Should an event occur that causes the switch that is receiving the BPDU to detect a loss of connection on its root port, such as in the event that a root port failure occurs, or a link failure occurs, for which the downstream switch is made immediately aware, the switch can instantly transition the alternate port.

This will begin the transition through the listening, learning and forwarding states and achieve recovery within a 2x forward delay period. In the event of any failure, where the link that provides a better path is reactivated, the spanning tree topology must again re-converge in order to apply the optimal spanning tree topology.

Topology Change MAC Instability



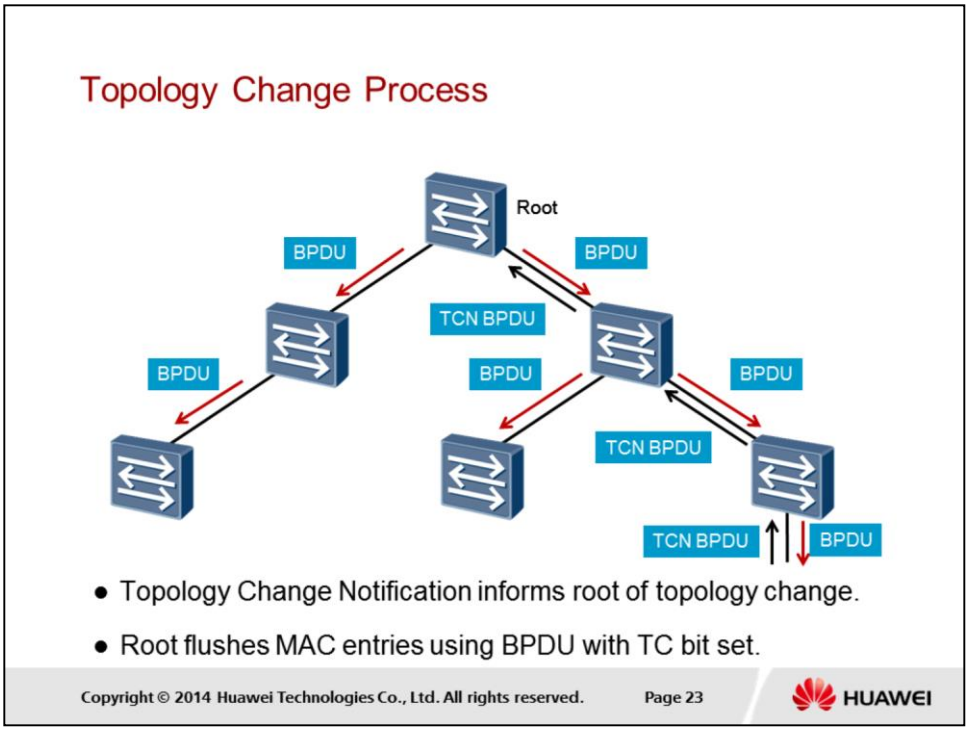
- Changes in the STP topology may invalidate MAC table entries.
- MAC table entries expire only after 300 seconds by default.

HUAWEI

In a converged spanning tree network, switches maintain filter databases, or MAC address tables to manage the propagation of frames through the spanning tree topology. The entries that provide an association between a MAC destination and the forwarding port interface are stored for a finite period of 300 seconds (5 minutes) by default. A change in the spanning tree topology however means that any existing MAC address table entries are likely to become invalid due to the alteration in the switching path, and therefore must be renewed.

The example demonstrates an existing spanning tree topology for which switch B has entries that allow Host A to be reached via interface Gigabit Ethernet 0/0/3 and Host B via interface Gigabit Ethernet 0/0/2. A failure is simulated on switch C for which the current root port has become inactive. This failure causes a recalculation of the spanning tree topology to begin and predictably the activation of the redundant link between switch C and switch B.

Following the re-convergence however, it is found that frames from Host A to Host B are failing to reach their destination. Since the MAC address table entries have yet to expire based on the 300 second rule, frames reaching switch B that are destined for Host B continue to be forwarded via port interface Gigabit Ethernet 0/0/2, and effectively become black holed as frames are forwarded towards the inactive port interface of switch C.

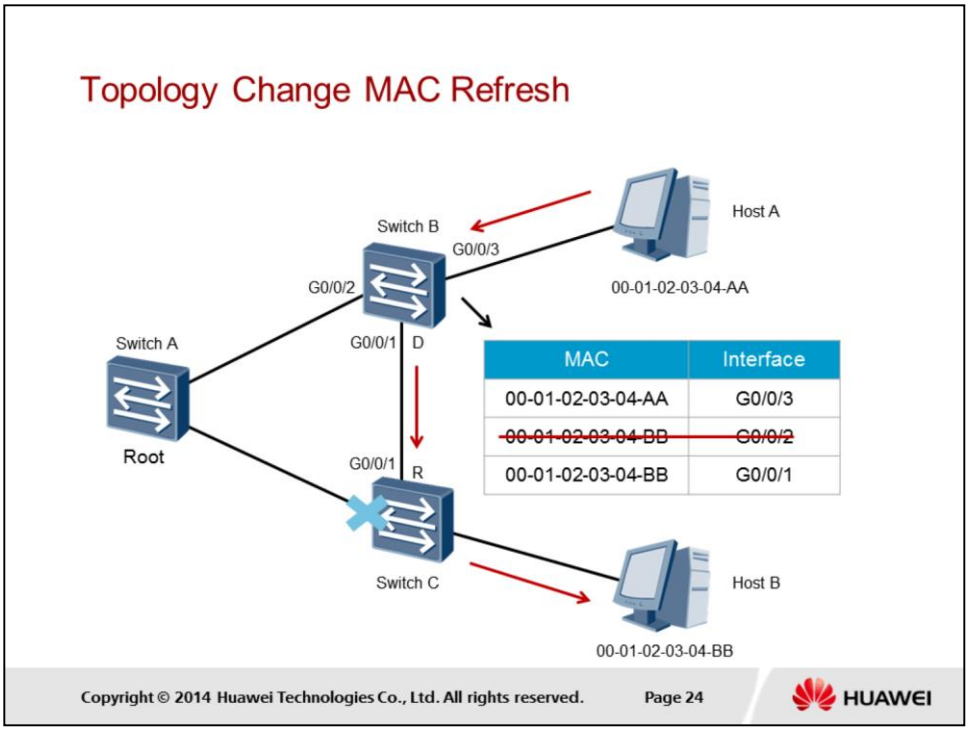


An additional mechanism must be introduced to handle the MAC entries timeout period issue that results in invalid path entries being maintained following spanning tree convergence. The process implemented is referred to as the Topology Change Notification (TCN) process, and introduces a new form of BPDU to the spanning tree protocol operation.

This new BPDU is referred to as the TCN BPDU and is distinguished from the original STP configuration BPDU through the setting of the BPDU type value to 128 (0x80). The function of the TCN BPDU is to inform the upstream root bridge of any change in the current topology, thereby allowing the root to send a notification within the configuration BPDU to all downstream switches, to reduce the timeout period for MAC address table entries to the equivalent of the forward delay timer, or 15 seconds by default.

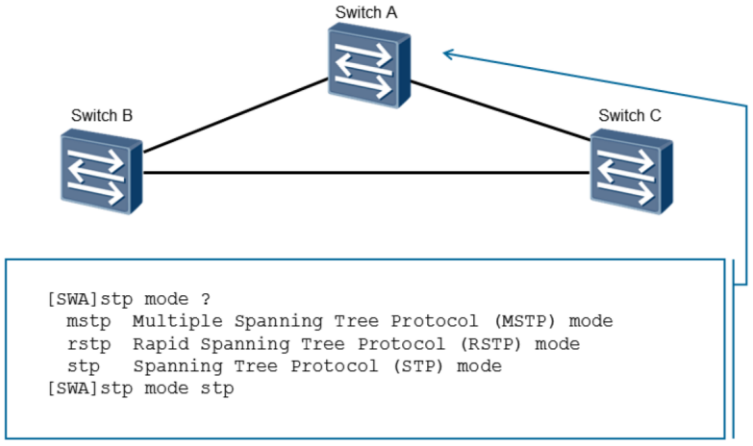
The flags field of the configuration BPDU contains two fields for Topology Change (TC) and Topology Change Acknowledgement (TCA). Upon receiving a TCN BPDU, the root bridge will generate a BPDU with both the TC and TCA bits set, to respectively notify of the topology change and to inform the downstream switches that the root bridge has received the TCN BPDU, and therefore transmission of the TCN BPDU should cease.

The TCA bit shall remain active for a period equal to the Hello timer (2 seconds), following which configuration BPDU generated by the root bridge will maintain only the TC bit for a duration of (MAX Age + forward delay), or 35 seconds by default.



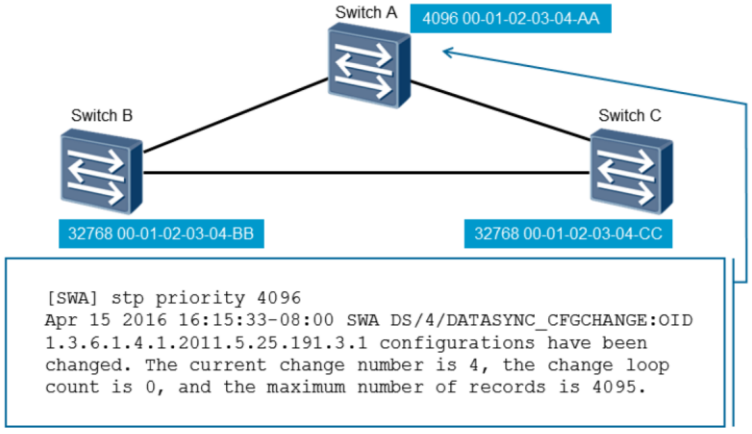
The effect of the TCN BPDU on the topology change process ensures that the root bridge is notified of any failure within the spanning tree topology, for which the root bridge is able to generate the necessary flags to flush the current MAC address table entries in each of the switches. The example demonstrates the results of the topology change process and the impact on the MAC address table. The entries pertaining to switch B have been flushed, and new updated entries have been discovered for which it is determined that Host B is now reachable via port interface Gigabit Ethernet 0/0/1.

STP Modes



Huawei Sx7 series switches to which the S5700 series model belongs, is capable of supporting three forms of spanning tree protocol. Using the *stp mode* command, a user is able to define the mode of STP that should be applied to an individual switch. The default STP mode for Sx7 series switches is MSTP, and therefore must be reconfigured before STP can be used.

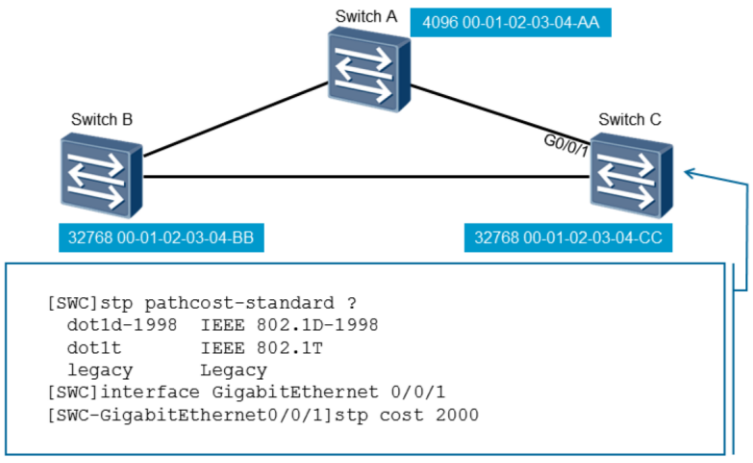
Assigning The Root



- Root can be set manually or by defining the switch as primary

As part of good switch design practice, it is recommended that the root bridge be manually defined. The positioning of the root bridge ensures that the optimal path flow of traffic within the enterprise network can be achieved through configuration of the bridge priority value for the spanning tree protocol. The *stp priority* [*priority*] command can be used to define the priority value, where *priority* refers to an integer value between 0 and 61440, assigned in increments of 4096. This allows for a total of 16 increments, with a default value of 32768. It is also possible to assign the root bridge for the spanning tree through the *stp root primary* command.

Assigning Path Cost

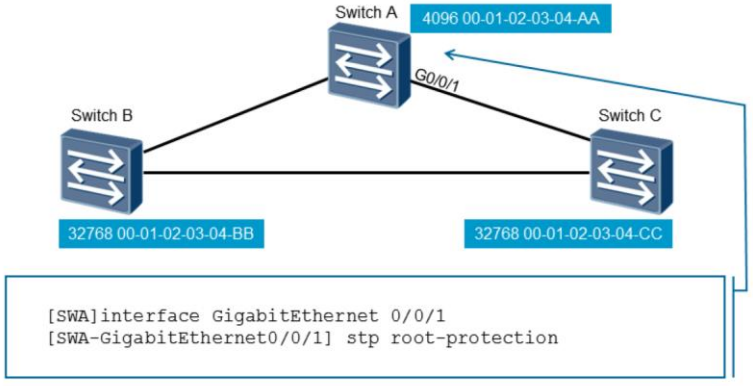


It has been understood that Huawei Sx7 series of switches support three forms of path cost standard in order to provide compatibility where required, however defaults to support the 802.1t path cost standard. The path cost standard can be adjusted for a given switch using the *stp pathcost-standard* { *dot1d-1998* | *dot1t* | *legacy* } command, where dot1d-1998, dot1t and legacy refer to the path cost standards described earlier in this section.

In addition, the path cost of each interface can also be assigned manually to support a means of detailed manipulation of the stp path cost. This method of path cost manipulation should be used with great care however as the path cost standards are designed to implement the optimal spanning tree topology for a given switching network and manipulation of the stp cost may result in the formation of a sub-optimal spanning tree topology.

The command *stp cost* [*cost*] is used, for which the cost value should follow the range defined by the path cost standard. If a Huawei legacy standard is used, the path cost ranges from 1 to 200000. If the IEEE 802.1D standard is used, the path cost ranges from 1 to 65535. If the IEEE 802.1t standard is used, the path cost ranges from 1 to 200000000.

Root Protection



- Root protection prevents changes to the topology as a result of root bridge transition, caused by receiving higher priority BPDU.

HUAWEI

If the root switch on a network is incorrectly configured or attacked, it may receive a BPDU with a higher priority and thus the root switch becomes a non-root switch, which causes a change of the network topology. As a result, traffic may be switched from high-speed links to low-speed links, causing network congestion.

To address this problem, the switch provides the root protection function. The root protection function protects the role of the root switch by retaining the role of the designated port. When the port receives a BPDU with a higher priority, the port stops forwarding packets and turns to the listening state, but it still retains a designated port role. If the port does not receive any BPDU with a higher priority for a certain period, the port status is restored from the listening state.

The configured root protection is valid only when the port is the designated port and the port maintains the role. If a port is configured as an edge port, or if a command known as loop protection is enabled on the port, root protection cannot be enabled on the port.

Configuration Validation

```
[SWA]display stp
-------[CIST Global Info][Mode STP]-------
CIST Bridge          :4096 .00-01-02-03-04-BB
Bridge Times         :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC       :4096 .00-01-02-03-04-BB / 0
CIST RegRoot/IRPC    :4096 .00-01-02-03-04-BB / 0
CIST RootPortId      :0.0
BPDU-Protection      :Disabled
TC or TCN received   :37
TC count per hello   :0
STP Converge Mode    :Normal
Share region-configuration :Enabled
Time since last TC   :0 days 0h:1m:29s
......
```

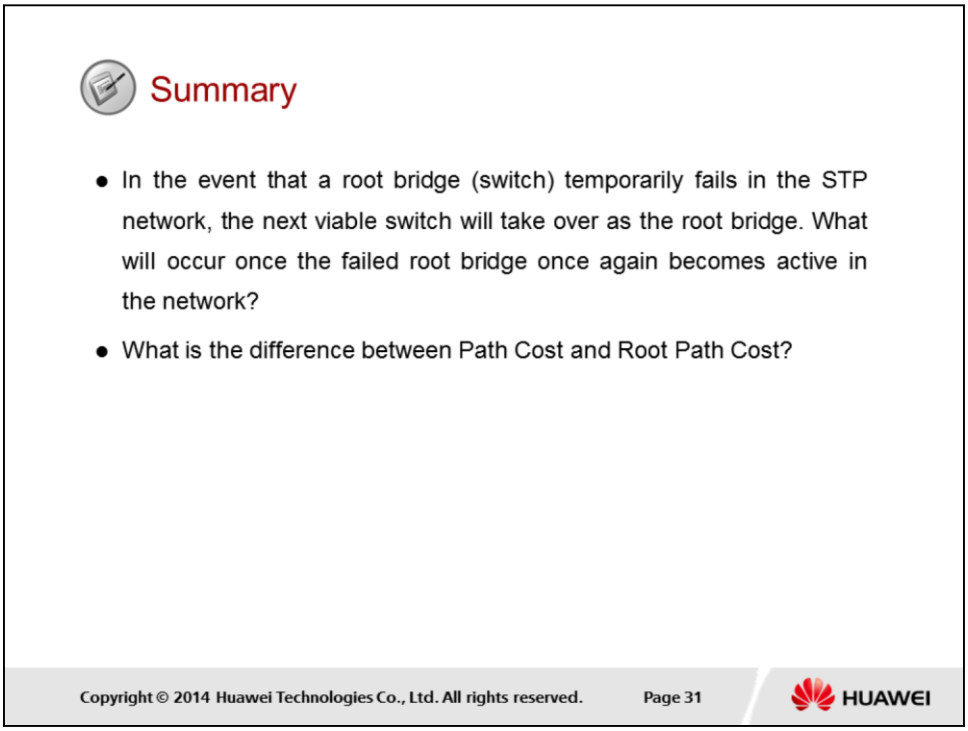

Using the *display stp* command, the current STP configuration can be determined. A number of timers exist for managing the spanning tree convergence, including the hello timer, max age timer, and forward delay, for which the values displayed represent the default timer settings, and are recommended to be maintained.

The current bridge ID can be identified for a given switch through the CIST Bridge configuration, comprised of the bridge ID and MAC address of the switch. Statistics provide information regarding whether the switch has experienced topology changes, primarily through the TC or TCN received value along with the last occurrence as shown in the time since last TC entry.

Configuration Validation

```
[SWA]display stp
……
----[Port1(GigabitEthernet0/0/1)][FORWARDING]----
 Port Protocol           :Enabled
 Port Role               :Designated Port
 Port Priority           :128
 Port Cost(Dot1T )       :Config=2000 / Active=2000
 Designated Bridge/Port  :4096.00-01-02-03-04-BB / 128.1
 Port Edged              :Config=default / Active=disabled
 Point-to-point          :Config=auto / Active=true
 Transit Limit           :147 packets/hello-time
 Protection Type         :Root
……
```


For individual interfaces on a switch it is possible to display this information via the *display stp* command to list all interfaces, or using the *display stp interface <interface>* command to define a specific interface. The state of the interface follows MSTP port states and therefore will display as either Discarding, Learning or Forwarding. Other valid information such as the port role and cost for the port are also displayed, along with any protection mechanisms applied.



1. Following the failure of the root bridge for a spanning tree network, the next best candidate will be elected as the root bridge. In the event that the original root bridge becomes active once again in the network, the process of election for the position of root bridge will occur once again. This effectively causes network downtime in the switching network as convergence proceeds.
2. The Root Path Cost is the cost associated with the path back to the root bridge, whereas the Path Cost refers to the cost value defined for an interface on a switch, which is added to the Root Path Cost, to define the Root Path Cost for the downstream switch.

Thank you

www.huawei.com

Rapid Spanning Tree Protocol

HUAWEI TECHNOLOGIES CO., LTD.

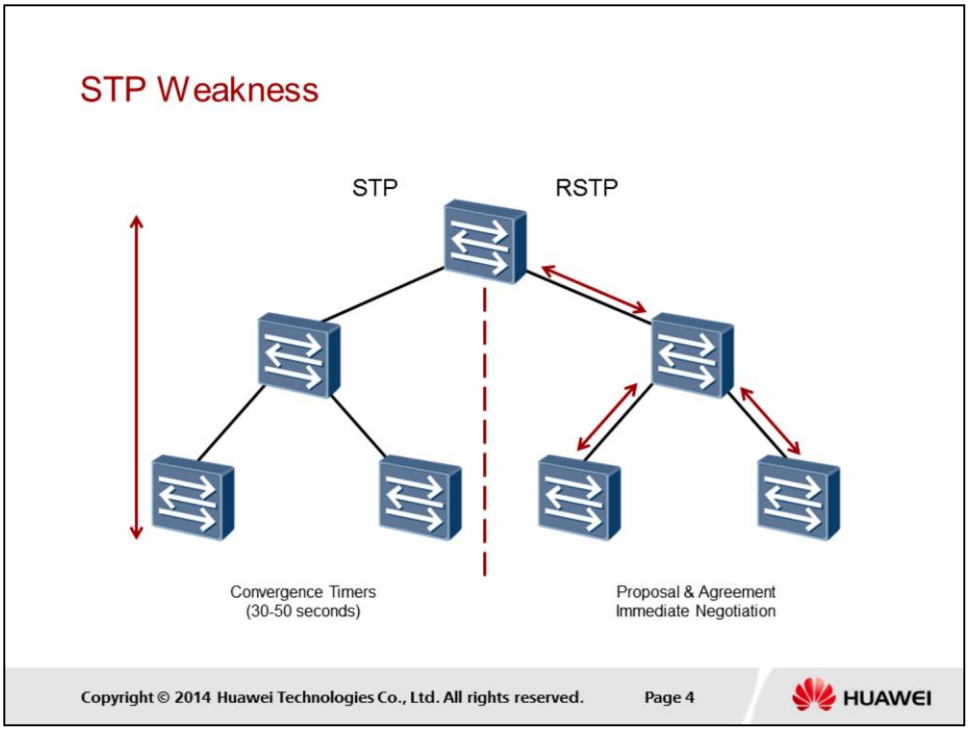
Foreword

The original STP standard was defined in 1998 for which a number of limitations were discovered, particularly in the time needed for convergence to occur. In light of this, Rapid Spanning Tree Protocol (RSTP) was introduced. The fundamental characteristics of RSTP are understood to follow the basis of STP, therefore the characteristic differences found within RSTP are emphasized within this section.

Objectives

Upon completion of this section, trainees will be able to:

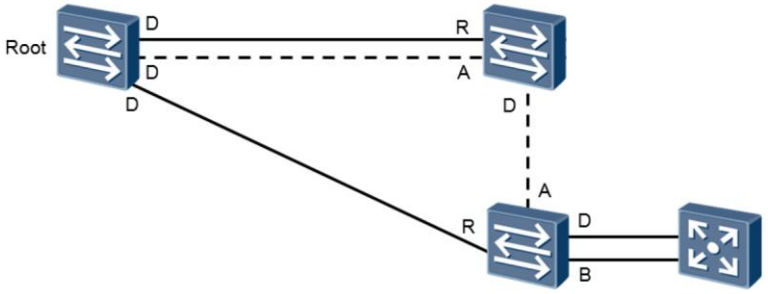
- Describe the characteristics associated with RSTP.
- Configure RSTP parameters.



STP ensures a loop-free network but has a slow network topology convergence speed, leading to service deterioration. If the network topology changes frequently, the connections on the STP capable network are frequently torn down, causing regular service interruption.

RSTP employs a proposal and agreement process which allows for immediate negotiation of links to take place, effectively removing the time taken for convergence based timers to expire before spanning tree convergence can occur. The proposal and agreement process tends to follow a cascading effect from the point of the root bridge through the switching network, as each downstream switch begins to learn of the true root bridge and the path via which the root bridge can be reached.

RSTP Port Roles

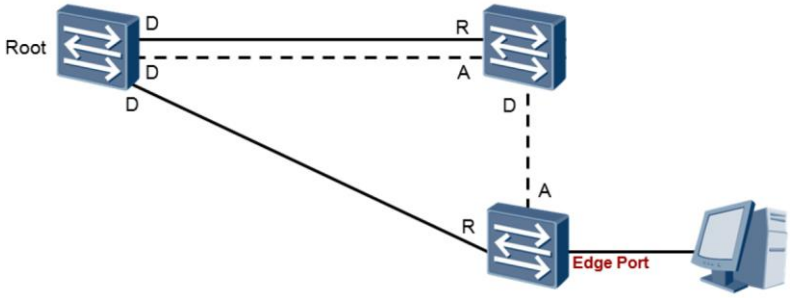


Roles	Description
Backup	A backup path to downstream nodes, where redundant links exist on the same LAN segment as the designated port.
Alternate	An alternate path to the root bridge that differs from the path provided by the root port of the switch.

Switches operating in RSTP mode implement two separate port roles for redundancy. The alternate port represents a redundant path to the root bridge in the event that the current path to the root bridge fails. The backup port role represents a backup for the path for the LAN segment in the direction leading away from the root bridge. It can be understood that a backup port represents a method for providing redundancy to the designated port role in a similar way that an alternate port provides a method of redundancy to the root port.

The backup port role is capable of existing where a switch has two or more connections to a shared media device such as that of a hub, or where a single point-to-point link is used to generate a physical loopback connection between ports on the same switch. In both instances however the principle of a backup port existing where two or more ports on a single switch connect to a single LAN segment still applies.

RSTP Edge Ports



- Systems that do not participate in RSTP connect to edge ports.
- Edge ports do not receive BDPU and can instantly forward data.

HUAWEI

In RSTP, a designated port on the network edge is called an edge port. An edge port directly connects to a terminal and does not connect to any other switching devices. An edge port does not receive configuration BPDU, so it does not participate in the RSTP calculation.

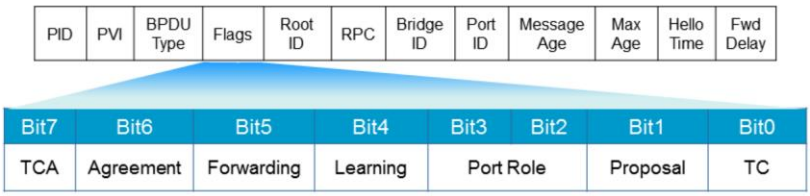
It can directly change from the Disabled state to the Forwarding state without any delay, just like an STP-incapable port. If an edge port receives bogus configuration BPDU from attackers, it is deprived of the edge port attributes and becomes a common STP port. The STP calculation is implemented again, causing network flapping.

Port States of RSTP

STP	RSTP	Port Role
Disabled	Discarding	Disabled
Blocking	Discarding	Alternate or Backup
Listening	Discarding	Root or Designated
Learning	Learning	Root or Designated
Forwarding	Forwarding	Root or Designated

RSTP introduces a change in port states that are simplified from five to three types. These port types are based on whether a port forwards user traffic and learns MAC addresses. If a port neither forwards user traffic nor learns MAC addresses, the port is in the Discarding state. The port is considered to be in a learning state where a port does not forward user traffic but learns MAC addresses. Finally where a port forwards user traffic and learns MAC addresses, the port is said to be in the Forwarding state.

RST BPDU



Port Role = 00 Unknown
01 Alternate/Backup Port
10 Root Port
11 Designated Port

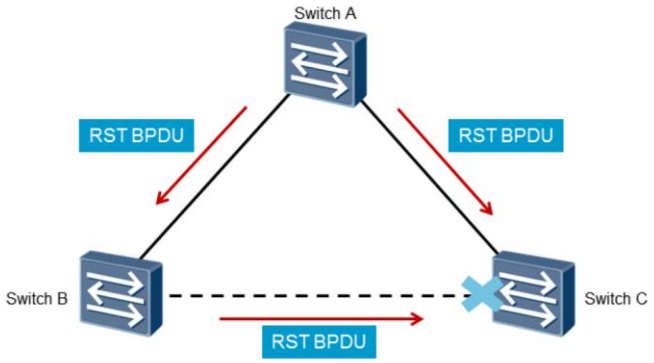
- Unused fields of the STP BPDU are active within RSTP.
- New capabilities are introduced as part of RSTP.

HUAWEI

The BPDU format employed in STP is also applied to RSTP with variance in some of the general parameters. In order to distinguish STP configuration BPDU from Rapid Spanning Tree BPDU, thus known as RST BPDU, the BPDU type is defined. STP defines a configuration BPDU type of 0 (0x00) and a Topology Change Notification BPDU (TCN BPDU) of 128 (0x80), RST BPDU are identified by the BPDU type value 2 (0x02). Within the flags field of the RST BPDU, additional parameter designations are assigned to the BPDU fields.

The flags field within STP implemented only the use of the Topology Change (TC) and Acknowledgement (TCA) parameters as part of the Topology Change process while other fields were reserved. The RST BPDU has adopted these fields to support new parameters. These include flags indicating the proposal and agreement process employed by RSTP for rapid convergence, the defining of the port role, and the port state.

RST BPDU

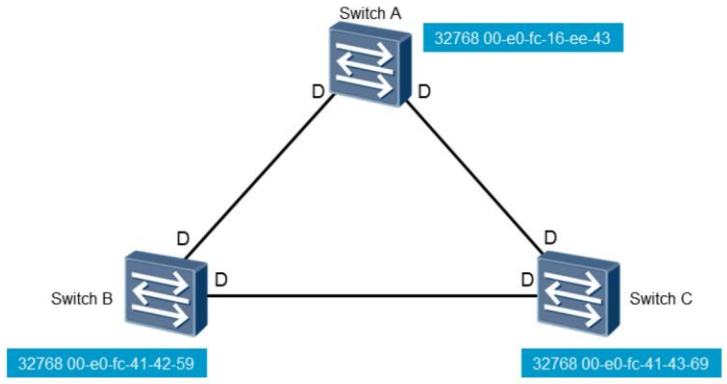


- Designated switches generate their own BDPU at Hello time, regardless of whether an RST BPDU has been received.

HUAWEI

In STP, after the topology becomes stable, the root bridge sends configuration BPDU at an interval set by the Hello timer. A non-root bridge does not send configuration BPDU until it receives configuration BPDU sent from the upstream device. This renders the STP calculation complicated and time-consuming. In RSTP, after the topology becomes stable, a non-root bridge sends configuration BPDU at Hello intervals, regardless of whether it has received the configuration BPDU sent from the root bridge; such operations are implemented on each device independently.

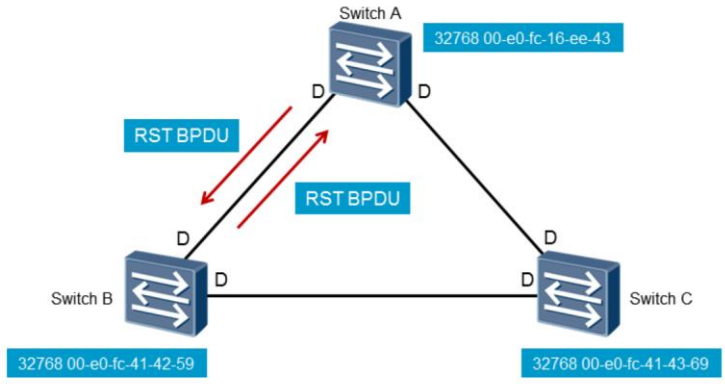
RSTP Convergence



- All RSTP enabled switches begin as root and send RST BPDU.
- Ports are set to a designated role and a discarding state.

The convergence of RSTP follows some of the basic principles of STP in determining initially that all switches upon initialization assert the role of root bridge, and as such assign each port interface with a designated port role. The port state however is set to a discarding state until such time as the peering switches are able to confirm the state of the link.

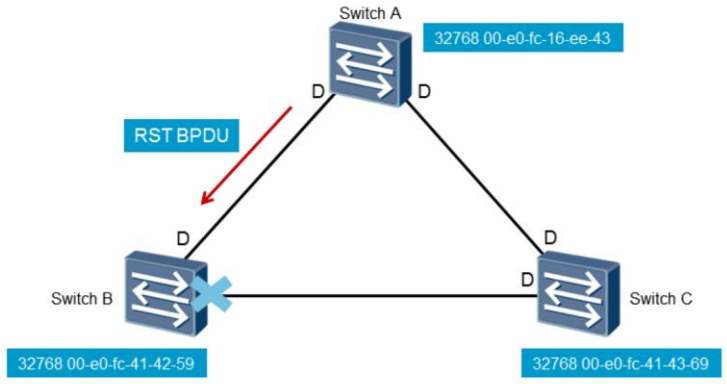
RST BPDU Proposal



- Proposals are sent in RST BPDU during root election.
- A switch will ignore a proposal if it has a better bridge ID

Each switch proclaiming to be the root bridge will negotiate the port states for a given LAN segment by generating an RST BPDU with the proposal bit set in the flags field. When a port receives an RST BPDU from the upstream designated bridge, the port compares the received RST BPDU with its own RST BPDU. If its own RST BPDU is superior to the received one, the port discards the received RST BPDU and immediately responds to the peering device with its own RST BPDU that includes a set proposal bit.

RSTP Synchronization Process



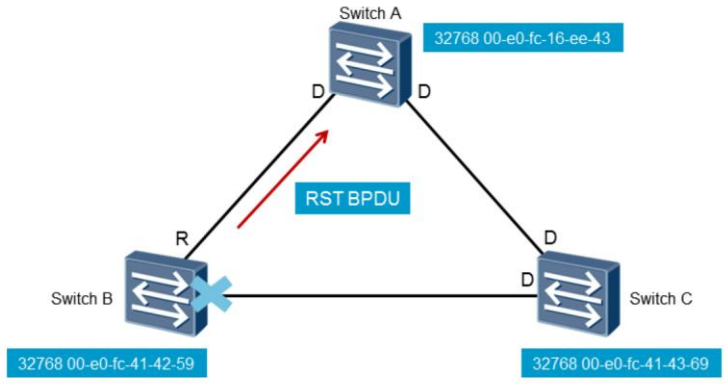
- Upon receiving a superior BPDU, Switch B will cease to send RST BDPU containing proposals and begin to synchronize.

HUAWEI

Since timers do not play a role in much of the RSTP topology convergence process as found with STP, it is important that the potential for switching loops during port role negotiation be restricted. This is managed by the implementation of a synchronization process that determines that following the receipt of a superior BPDU containing the proposal bit, the receiving switch must set all downstream designated ports to discarding as part of the sync process.

Where the downstream port is an alternate port or an edge port however, the status of the port role remains unchanged. The example demonstrates the temporary transition of the designated port on the downstream LAN segment to a discarding state, and therefore blocking any frame forwarding during the upstream proposal and agreement process.

RST BPDU Agreement

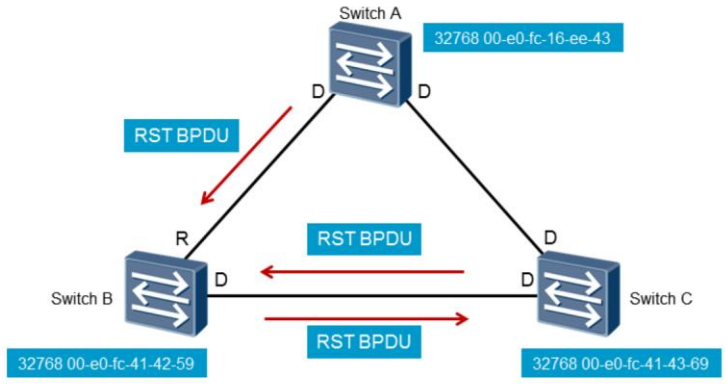


- Once all downstream non-edge designated ports have been blocked, Switch B will send an agreement with the RST BPDU

HUAWEI

The confirmed transition of the downstream designated port to a discarding state allows for an RST BPDU to be sent in response to the proposal sent by the upstream switch. During this stage the port role of the interface has been determined to be the root port and therefore the agreement flag and port role of root are set in the flags field of the RST BPDU that is returned in response to the proposal.

RSTP Converged Link

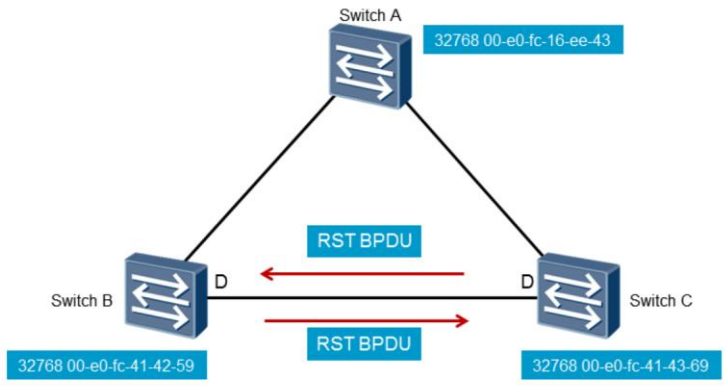


- The downstream port is again unblocked and a new round of synchronization occurs between Switch B and Switch C.

HUAWEI

During the final stage of the proposal and agreement process, the RST BPDU containing the agreement bit is received by the upstream switch, allowing the designated port to transition immediately from a discarding state to forwarding state. Following this, the downstream LAN segment(s) will begin to negotiate the port roles of the interfaces using the same proposal and agreement process.

Link/Root Failure

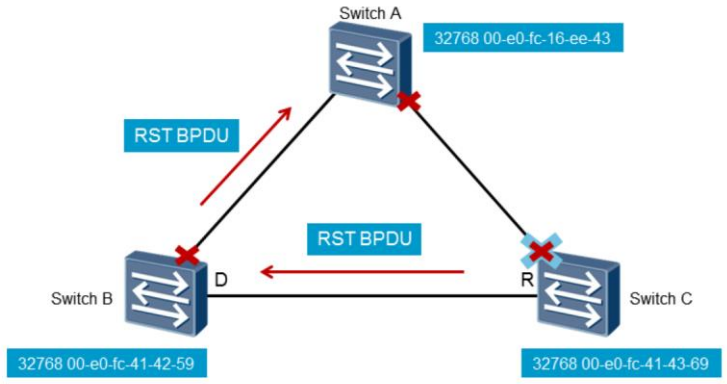


- Loss of upstream RST BPDU signals link/device failure.
- Proposal and agreement based convergence will ensue.

HUAWEI

In STP, a device has to wait a Max Age period before determining a negotiation failure. In RSTP, if a port does not receive configuration BPDUs sent from the upstream device for three consecutive Hello intervals, the communication between the local device and its peer fails, causing the proposal and agreement process to be initialized in order to discover the port roles for the LAN segment.

Topology Change Process



- During the sending of an agreement, addresses are flushed for all ports except the port on which the RST BPDU was received.

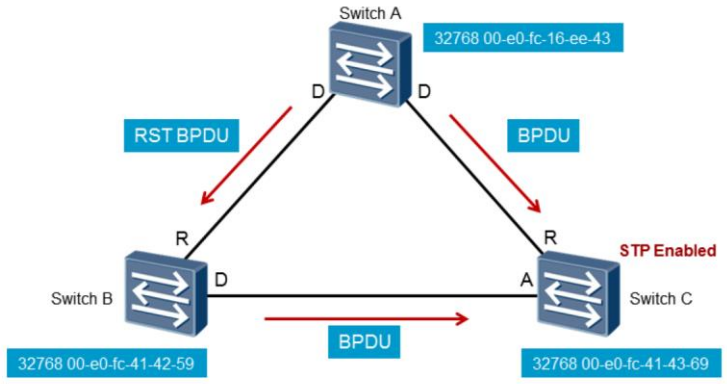
HUAWEI

Topology changes affect RSTP similarly to the way STP is affected, however there are some minor differences between the two. In the example, a failure of the link has occurred on switch C. Switch A and switch C will detect the link failure immediately and flush the address entries for ports connected to that link. An RST BPDU will begin to negotiate the port states as part of the proposal and agreement process, following which a Topology Change notification will occur, together with the forwarding of the RST BPDU containing the agreement.

This RST BPDU will have both the Agreement bit and also the TC bit set to 1, to inform upstream switches of the need to flush their MAC entries on all port interfaces except the port interface on which the RST BPDU containing the set TC bit was received.

The TC bit will be set in the periodically sent RST BPDU, and forwarded upstream for a period equivalent to Hello Time+1 second, during which all relevant interfaces will be flushed and shall proceed to re-populate MAC entries based on the new RSTP topology. The red (darker) ‘x’ in the example highlights which interfaces will be flushed as a result of the topology change.

STP Inter-Operation

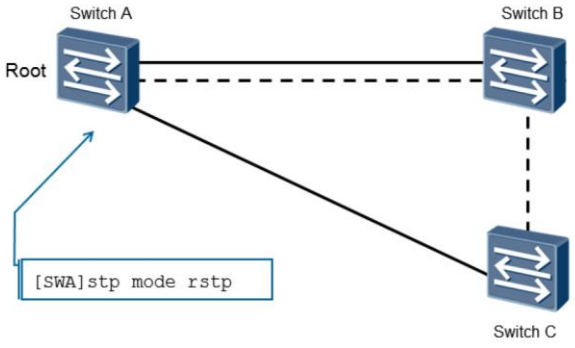


- RSTP switch ports will revert to STP when connected to a LAN segment containing an STP enabled device.

HUAWEI

The implementation of STP within an RSTP based switching topology is possible, however is not recommended since any limitation pertaining to STP becomes apparent within the communication range of the STP enabled switch. A port involved in the negotiation process for establishing its role within STP must wait for a period of up to 50 seconds before convergence can be completed, as such the benefits of RSTP are lost.

Setting the Mode



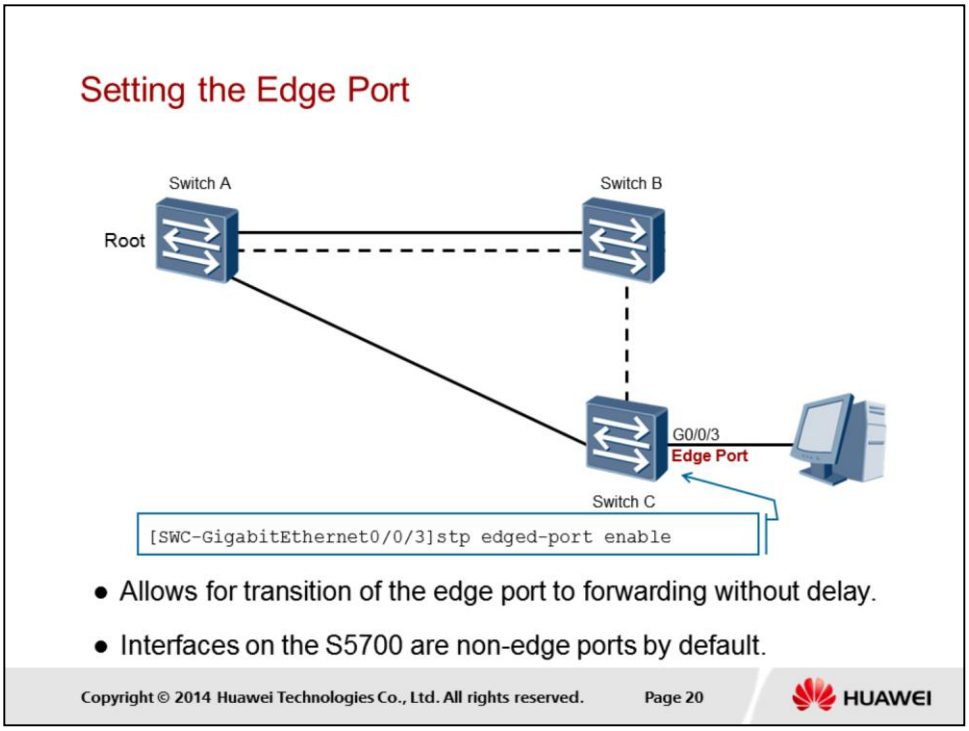
- The *stp mode rstp* command allows all ports of the switch to generate RST BPDU.

The configuration of the spanning tree mode of Sx7 switches requires that the *stp mode* command be used to set the mode to RSTP. In doing so the Sx7 series switch will generate RST BPDU in relation to RSTP, as opposed to other spanning tree implementations. This command is configured from the system-view and should be applied to all switches participating in the rapid spanning tree topology.

Configuration Validation

```
[SWA]display stp
-------[CIST Global Info][Mode RSTP]-------
CIST Bridge          :32768.00-e0-fc-16-ee-43
Bridge Times         :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC       :32768.00-e0-fc-16-ee-43 / 0
CIST RegRoot/IRPC    :32768.00-e0-fc-16-ee-43 / 0
CIST RootPortId      :0.0
BPDU-Protection      :Disabled
TC or TCN received   :37
TC count per hello   :0
STP Converge Mode    :Normal
Share region-configuration :Enabled
Time since last TC   :0 days 0h:14m:43s
```

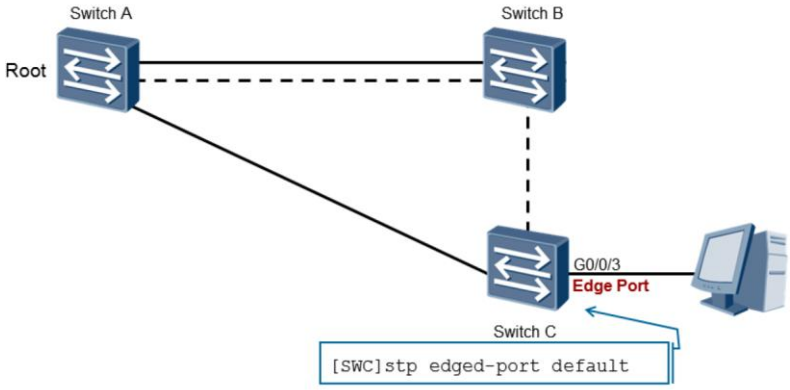

The *display stp* command will provide relative information regarding RSTP configuration as many of the parameters follow the principle STP architecture. The mode information will determine as to whether a switch is currently operating using RSTP.



An edge interface defines a port that does not participate in the spanning tree topology. These interfaces are used by end systems to connect to the switching network for the purpose of forwarding frames. Since such end systems do not require to negotiate port interface status, it is preferable that the port be transitioned directly to a forwarding state to allow frames to be forwarded over this interface immediately.

The *stp edged-port enable* command is used to switch a port to become an edge port, as all ports are considered non-edge ports on a switch by default. In order to disable the edge port the *stp edged-port disable* command is used. These commands apply only to a single port interface on a given switch. It is important to note that the edge port behavior is associated with RSTP as defined in the IEEE 802.1D-2004 standards documentation, however due to the VRP specific application of the underlying RSTP state machine to STP (which also results in the RSTP port states being present in STP), it is also possible to apply the RSTP edge port settings to STP within Huawei Sx7 series products.

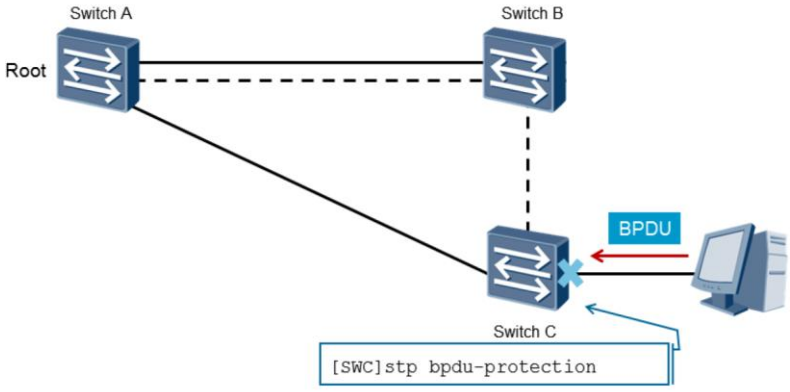
Setting the Edge Port



- All ports on the switch will be configured as edge ports.
- Care should be taken with this command to avoid STP loops.

In the event that multiple ports on a switch are to be configured as edge ports, the *stp edged-port default* command is applied which enforces that all port interfaces on a switch become edge ports. It is important to run the stp edged-port disable command on the ports that need to participate in STP calculation between devices, so as to avoid any possible loops that may be caused as a result of STP topology calculations.

BPDU Protection



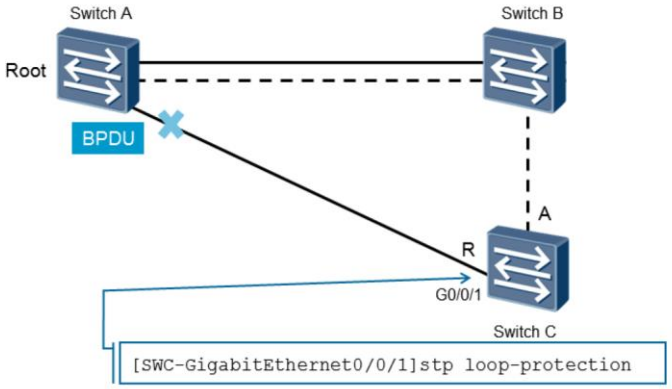
- BPDU protection prevents the malicious injection of BPDU into RSTP.

HUAWEI

The port that is directly connected to a user terminal such as a PC or a file server, is understood to be configured as an edge port to ensure fast transition of the port status. Usually, no BPDU are sent to edge ports, however if the switch is attacked by pseudo BPDU, the switch sets edge ports as non-edge ports. After these edge ports receive a BPDU the spanning tree topology is recalculated, and as a result network flapping occurs.

To defend against pseudo BPDU attacks, RSTP provides BPDU protection. After BPDU protection is enabled, the switch shuts down the edge port that receives BPDU and informs any active network management station (NMS). The edge ports that are shut down by the switch can be manually started only by the network administrator. The *stp bpdu-protection* command should be used to enable bpdu protection and is configured globally within the system-view.

Loop Protection



- If BDPU fail to be received by the downstream switch, the root port is blocked to prevent switching loops from occurring.

HUAWEI

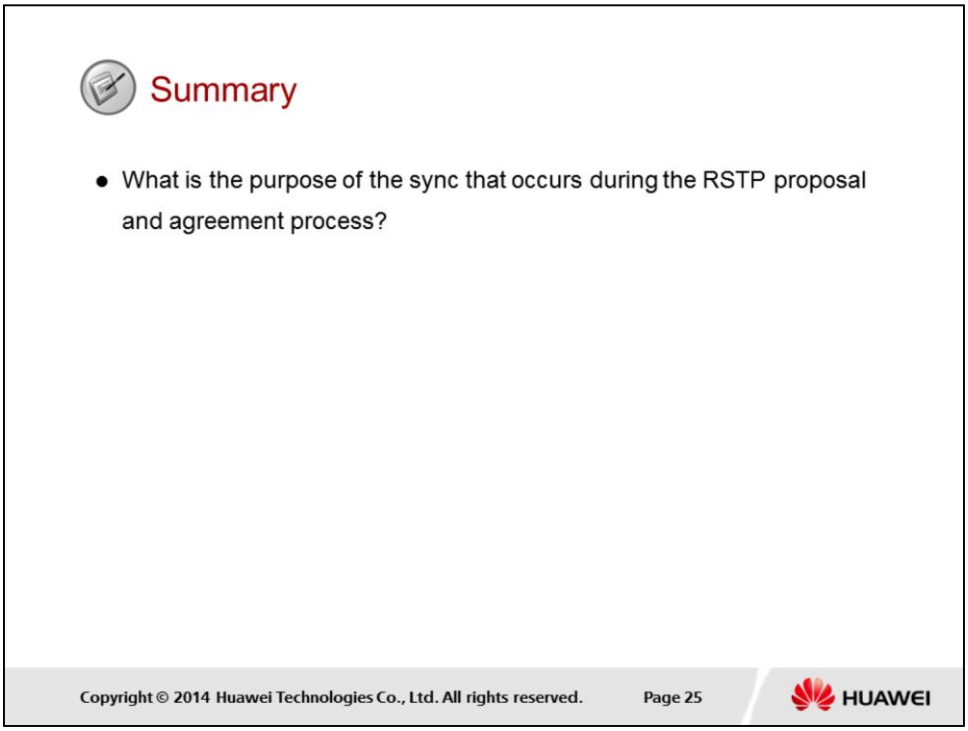
The switch maintains the status of the root port and blocked ports by continually receiving BPDU from the upstream switch. If the root switch cannot receive BPDU from the upstream switch due to link congestion or unidirectional link failure, the switch re-selects a root port. The previous root port then becomes a designated port and the blocked ports change to the forwarding state. As a result, loops may occur on the network.

The switch provides loop protection to prevent network loops. After the loop protection function is enabled, the root port is blocked if it cannot receive BPDU from the upstream switch. The blocked port remains in the blocked state and does not forward packets. This prevents loops on the network. If an interface is configured as an edge interface or root protection is enabled on the interface, loop protection cannot be enabled on the interface. The *stp loop-protection* command should be applied to enable this feature in the interface-view.

Configuration Validation

```
[SWC]display stp interface GigabitEthernet 0/0/1
----[CIST][Port1(GigabitEthernet0/0/1)][FORWARDING]----
 Port Protocol           :Enabled
 Port Role               :Root Port
 Port Priority           :128
 Port Cost(Dot1T )       :Config=auto / Active=20000
 Designated Bridge/Port  :32768.00-e0-fc-16-ee-43 / 128.1
 Port Edged              :Config=default / Active=disabled
 Point-to-point          :Config=auto / Active=true
 Transit Limit           :147 packets/hello-time
 Protection Type         :Loop
 Port STP Mode           :RSTP
 Port Protocol Type      :Config=auto / Active=dot1s
 BPDU Encapsulation      :Config=stp / Active=stp
......
```


Validation of the RSTP configuration for a given interface is attained through the *display stp interface <interface>* command. The associated information will identify the port state of the interface as either Discarding, Learning or Forwarding. Relevant information for the port interface including the port priority, port cost, the port status as an edge port or supporting point-to-point etc, are defined.



1. The sync is a stage in the convergence process that involves the blocking of designated ports while RST BPDU are transmitted containing proposal and agreement messages to converge the switch segment. The process is designed to ensure that all interfaces are in agreement as to their port roles in order to ensure that no switching loops will occur once the designated port to any downstream switch is unblocked.

Thank you

www.huawei.com

Segmenting the IP Network

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

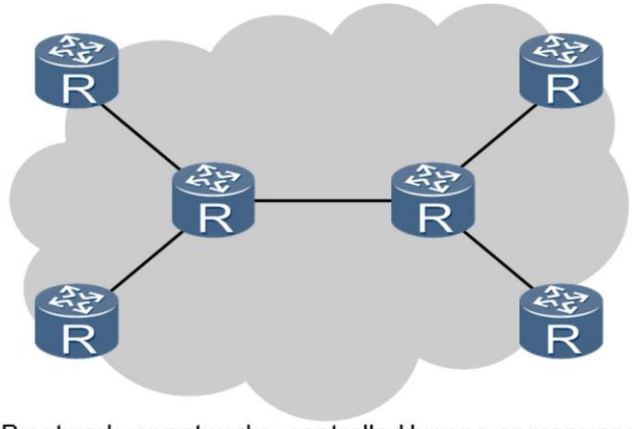
The forwarding of frames and switching has introduced the data link layer operations, and in particular the role of IEEE 802 based standards as the supporting underlying communication mechanism, over which upper layer protocol suites generally operate. With the introduction of routing, the physics that define upper layer protocols and internetwork communication are established. An enterprise network domain generally consists of multiple networks for which routing decisions are needed to ensure optimal routes are used, in order to forward IP packets (or datagrams) to intended network destinations. This section introduces the foundations on which such IP routing is based.

Objectives

Upon completion of this section, trainees will be able to:

- Explain the principles that govern IP routing decisions.
- Explain the basic requirements for packet forwarding.

Autonomous Systems



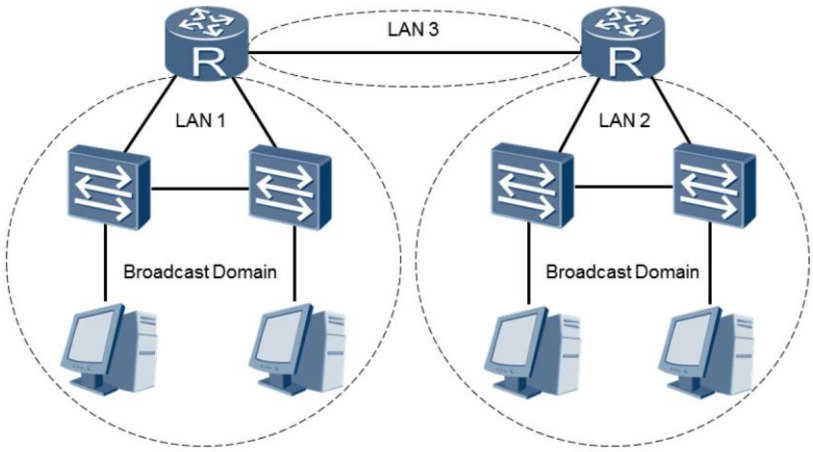
- An IP network, or networks, controlled by one or more operators with a clear policy that governs how routing decisions are made.

HUAWEI

An enterprise network generally can be understood as an instance of an autonomous system. As defined within RFC 1030, an autonomous system or AS, as it is also commonly known, is a connected group of one or more IP prefixes run by one or more network operators which has a SINGLE and CLEARLY DEFINED routing policy.

The concept of autonomous systems originally considered the existence of a single routing protocol, however as networks have evolved, it is possible to support multiple routing protocols that interoperate through the injection of routes from one protocol to another. A routing policy can be understood to be a set of rules that determine how traffic is managed within an autonomous system, to which a single, or multiple operator(s) must adhere to.

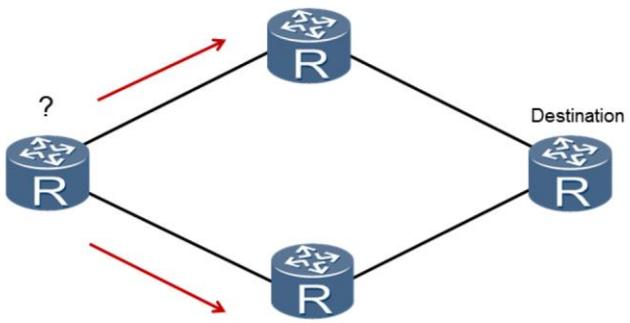
Local Area Network and Broadcast Domains



HUAWEI

The principles surrounding switching have dealt mainly with the forwarding of traffic within the scope of a local area network and the gateway, which has until now defined the boundary of the broadcast domain. Routers are the primary form of network layer device used to define the gateway of each local area network and enable IP network segmentation. Routers generally function as a means for routing packets from one local network to the next, relying on IP addressing to define the IP network to which packets are destined.

Routing Decisions



- Routers are responsible for the decision making process that determines the path via which packets are forwarded.

HUAWEI

The router is responsible for determining the forwarding path via which packets are to be sent en route to a given destination. It is the responsibility of each router to make decisions as to how the data is forwarded. Where a router has multiple paths to a given destination, route decisions based on calculations are made to determine the best next hop to the intended destination. The decisions governing the route that should be taken can vary depending on the routing protocol in use, ultimately relying on metrics of each protocol to make decisions in relation to varying factors such as bandwidth and hop count.

IP Routing Table

```
[Huawei]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 2        Routes : 2

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        127.0.0.0/8     Direct  0    0           D   127.0.0.1       InLoopBack0
        127.0.0.1/32    Direct  0    0           D   127.0.0.1       InLoopBack0
```

- The IP routing table lists the networks that are reachable via the router. Packets that have no route are subsequently discarded.

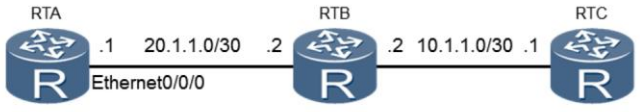
  HUAWEI

Routers forward packets based on routing tables and a forwarding information base (FIB), and maintain at least one routing table and one FIB. Routers select routes based on routing tables and forward packets based on the FIB. A router uses a local routing table to store protocol routes and preferred routes. The router then sends the preferred routes to the FIB to guide packet forwarding. The router selects routes according to the priorities of protocols and costs stored in the routing table. A routing table contains key data for each IP packet.

The destination & mask are used in combination to identify the destination IP address or the destination network segment where the destination host or router resides. The protocol (Proto) field, indicates the protocol through which routes are learned. The preference (Pre) specifies the preference value that is associated with the protocol, and is used to decide which protocol is applied to the routing table where two protocols offer similar routes. The router selects the route with the highest preference (the smallest value) as the optimal route.

A cost value represents the metric that is used to distinguish when multiple routes to the same destination have the same preference, the route with the lowest cost is selected as the optimal route. A next hop value indicates the IP address of the next network layer device or gateway that an IP packet passes through. In the example given a next hop of 127.0.0.1 refers to the local interface of the device as being the next hop. Finally the interface parameter indicates the outgoing interface through which an IP packet is forwarded.

Routing Decisions – Longest Match




```
[RTA]display ip routing-table
Destination/Mask  Proto   Pre  Cost      Flags NextHop         Interface
10.1.1.0/24       Static  60   0           RD  20.1.1.2        Ethernet0/0/0
10.1.1.0/30       Static  60   0           RD  20.1.1.2        Ethernet0/0/0
```

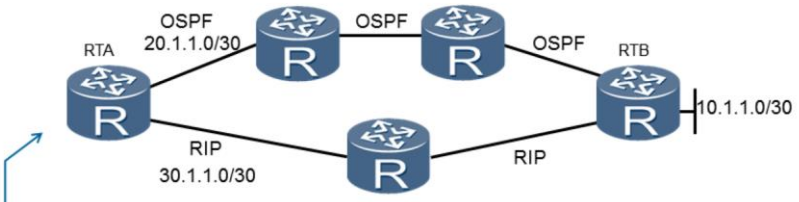
- Routes to the same network destination will be initially compared and chosen based on a longest match.

In order to allow packets to reach their intended destination, routers must make specific decisions regarding the routes that are learned and which of those routes are applied. A router is likely to learn about the path to a given network destination via routing information that is advertized from neighboring routers, alternatively it is possible for the statically applied routes to be manually implemented through administrator intervention.

Each entry in the FIB table contains the physical or logical interface through which a packet is sent in order to reach the next router. An entry also indicates whether the packet can be sent directly to a destination host in a directly connected network. The router performs an "AND" operation on the destination address in the packet and the network mask of each entry in the FIB table.

The router then compares the result of the "AND" operation with the entries in the FIB table to find a match. The router chooses the optimal route to forward packets according to the best or "longest" match. In the example, two entries to the network 10.1.1.0 exist with a next hop of 20.1.1.2. Forwarding to the destination of 10.1.1.1 will result in the longest match principle being applied, for which the network address 10.1.1.0/30 provides the longest match.

Routing Decisions – Preference




```
[RTA]display ip routing-table
Destination/Mask  Proto   Pre  Cost  Flags NextHop         Interface
10.1.1.0/30       OSPF    10   60    RD    20.1.1.2        Ethernet0/0/0
......
```

Route	Direct	OSPF	Static	RIP
Preference	0	10	60	100

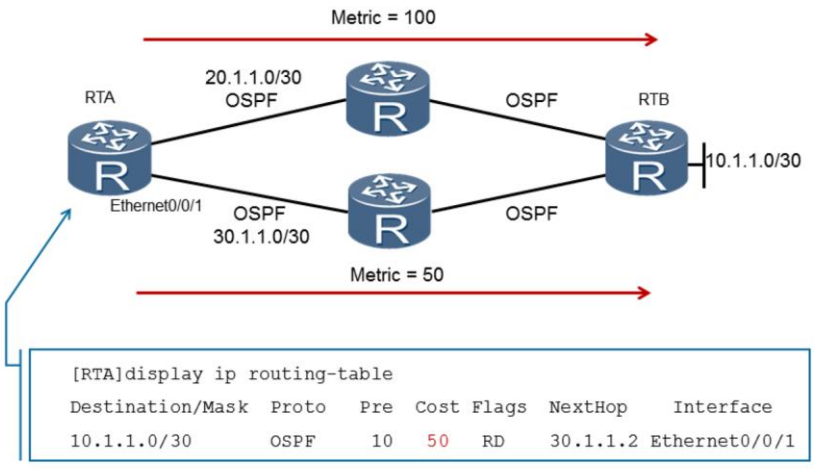
HUAWEI

A routing table may contain the routes originating from multiple protocols to a given destination. Not all routing protocols are considered equal, and where the longest match for multiple routes of differing routing protocols to the same destination are equal, a decision must be made regarding which routing protocol (including static routes) will take precedence.

Only one routing protocol at any one time determines the optimal route to a destination. To select the optimal route, each routing protocol (including the static route) is configured with a preference (the smaller the value, the higher the preference). When multiple routing information sources coexist, the route with the highest preference is selected as the optimal route and added to the local routing table.

In the example, two protocols are defined that provide a means of discovery of the 10.1.1.0 network via two different paths. The path defined by the RIP protocol appears to provide a more direct route to the intended destination, however due to the preference value, the route defined by the OSPF protocol is preferred and therefore installed in the routing table as the preferred route. A summary of the default preference values of some common routing mechanisms are provided to give an understanding of the default preference order.

Routing Decisions – Metric

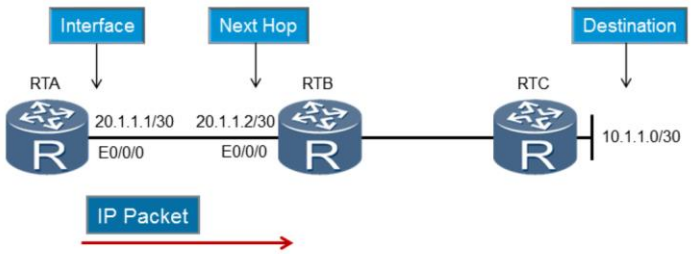



```
[RTA]display ip routing-table
Destination/Mask  Proto   Pre  Cost      Flags NextHop         Interface
10.1.1.0/30       OSPF    10   50        RD    30.1.1.2        Ethernet0/0/1
```


Where the route is unable to be distinguished by either a longest match value or preference, the cost metric is taken as the decision maker in identifying the route that should be installed in the routing table. Cost represents the length of a path to a destination network.

Each segment provides a cost metric value along a path that is combined to identify the cost of the route. Another common factor is network bandwidth, on which the cost mechanism is sometimes based. A link with a higher speed (capacity) represents a lower cost value, allowing preference of one path over another to be made, whilst links of equal speed are given a balanced cost for efficient load balancing purposes. A lower metric always takes precedence and therefore the metric of 50 as shown in the example, defines the optimal route to the given destination for which an entry can be found in the routing table.

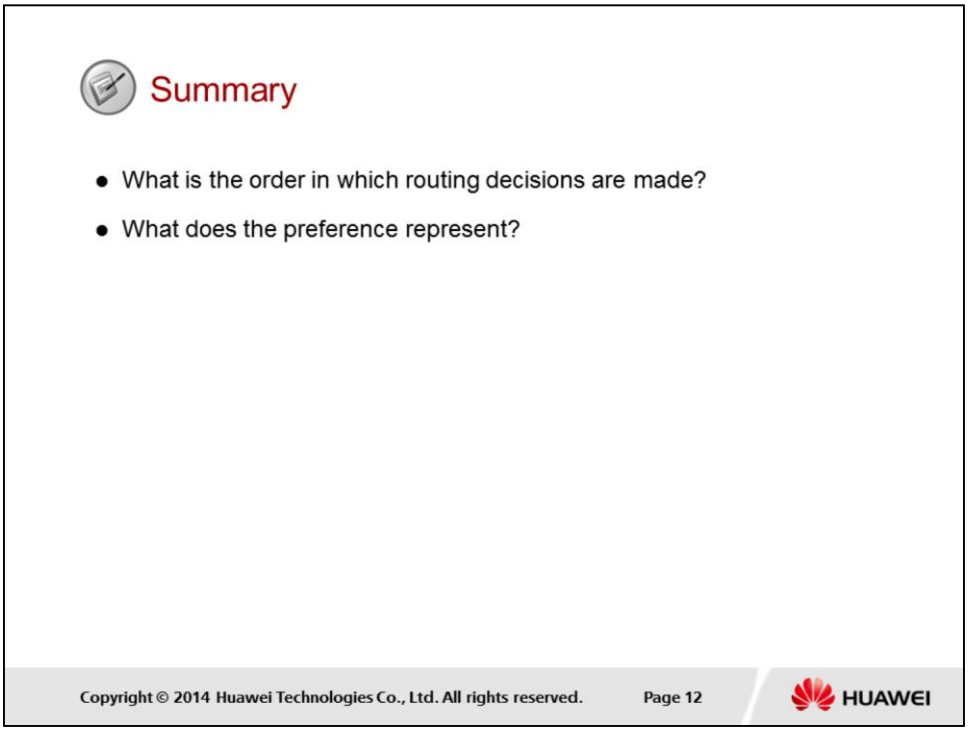
Routing Table Forwarding Requirements



- The forwarding of packets requires that the destination be known as well as the forwarding interface and next hop.

HUAWEI

The capability of a router to forward an IP packet to a given destination requires that certain forwarding information be known. Any router wishing to forward an IP packet must firstly be aware of a valid destination address to which the packet is to be forwarded, this means that an entry must exist in the routing table that the router is able to consult. This entry must also identify the interface via which IP packets must be transmitted and the next hop along the path, to which the packet is expected to be received before consultation for the next forwarding decision is performed.



1. Routing decisions are made initially based on the longest match value, regardless of the preference value assigned for routes to the same network. If the longest match value for two routes to the same destination is equal, the preference shall be used, where the preference is also equal, the metric shall be used. In cases where the metric value is also the same, protocols will commonly apply a form of load balancing of data over the equal cost links.
2. The preference is typically used to denote the reliability of a route over routes that may be considered less reliable. Vendors of routing equipment may however assign different preference values for protocols that are supported within each vendors own product. The preference values of some common routing protocols supported by Huawei routing devices can be found within this section.

Thank you

www.huawei.com

IP Static Routes

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

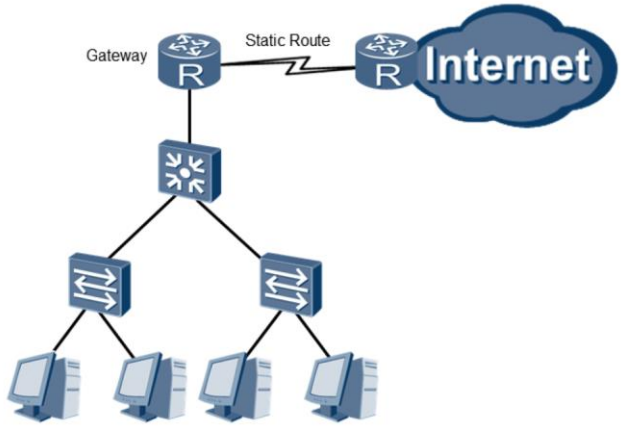
The implementation of routes within the IP routing table of a router can be defined manually using static routes or through the use of dynamic routing protocols. The manual configuration of routes enables direct control over the routing table, however may result in route failure should a router's next hop fail. The configuration of static routes however is often used to compliment dynamic routing protocols to provide alternative routes in the event dynamically discovered routes fail to provide a valid next hop. Knowledge of the various applications of static routes and configuration is necessary for effective network administration.

Objectives

Upon completion of this section, trainees will be able to:

- Explain the different applications for static routes.
- Successfully configure static routes in the IP routing table.

Application for Static Route

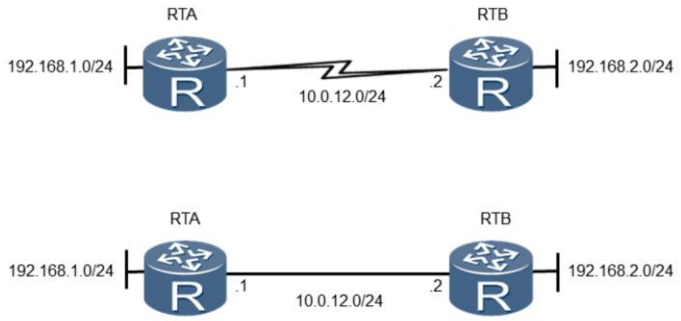


- Static routes define a means of path selection to other networks.

HUAWEI

A static route is a special route that is manually configured by a network administrator. The disadvantage of static routes is that they cannot adapt to the change in a network automatically, so network changes require manual reconfiguration. Static routes are fit for networks with comparatively simple structures. It is not advisable to configure and maintain static routes for a network with a complex structure. Static routes do however reduce the effect of bandwidth and CPU resource consumption that occurs when other protocols are implemented.

Static Route Behavior



- The forwarding of packets based on a serial interface requires that the outbound interface be defined.

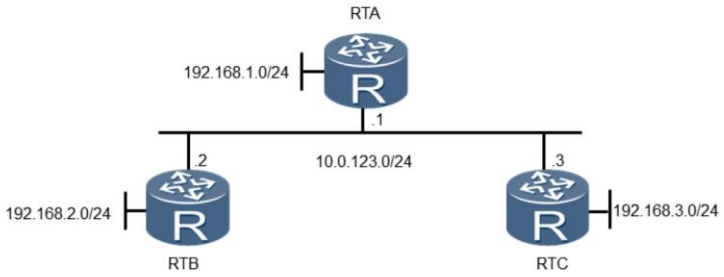
HUAWEI

Static routes can be applied to networks that use both serial and Ethernet based media, however in each situation the conditions of applying the static route vary in which either the outbound interface or the next-hop IP address must be defined.

The serial medium represents a form of point-to-point (P2P) interface for which the outbound interface must be configured. For a P2P interface, the next-hop address is specified after the outbound interface is specified. That is, the address of the remote interface (interface on the peer device) connected to this interface is the next-hop address.

For example, the protocol used to encapsulate over the serial medium is the Point-to-Point protocol (PPP). The remote IP address is obtained following PPP negotiation, therefore it is necessary to specify only the outbound interface. The example also defines a form of point-to-point Ethernet connection, however Ethernet represents a broadcast technology in nature and therefore the principles of point-to-point technology do not apply.

Static Route Behavior



- The forwarding of packets over broadcast networks such as Ethernet, requires that the next hop be defined.

In the case of broadcast interfaces such as Ethernet, the next hop must be defined. Where the Ethernet interface is specified as the outbound interface, multiple next hops are likely to exist and the system will not be able to decide which next hop is to be used. In determining the next hop, a router is able to identify the local connection over which the packet should be received.

In the example, packets intended for the destination of 192.168.2.0/24 should be forwarded to the next hop of 10.0.123.2 to ensure delivery. Alternatively reaching the destination of 192.168.3.0 requires that the next hop of 10.0.123.3 be defined.

Configuring a Static Route

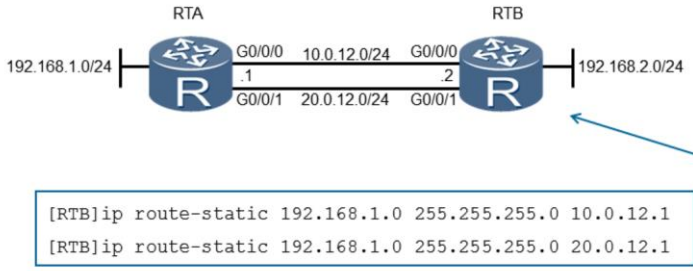

```
[RTB]ip route-static 192.168.1.0 255.255.255.0 10.0.12.1
[RTB]ip route-static 192.168.1.0 255.255.255.0 Serial 1/0/0
[RTB]ip route-static 192.168.1.0 24 Serial 1/0/0
```

- A static route can be configured based on one of three variations.

HUAWEI

The configuration of the static route is achieved using the *ip route-static ip-address* { *mask* | *mask-length* } *interface-type interface-number* [*nexthop-address*] where the *ip-address* refers to the network or host destination address. The mask field can be defined as either a mask value or based on the prefix number. In the case of a broadcast medium such as Ethernet, the next hop address is used. Where a serial medium is used, the interface-type and interface-number are assigned (e.g. serial 1/0/0) to the command to define the outgoing interface.

Static Route Load Balancing



- Static routes support load balancing to the same destination where the cost of routes are equal.

HUAWEI

Where equal cost paths exist between the source and destination networks, load balancing can be implemented to allow traffic to be carried over both links. In order to achieve this using static routes, both routes must meet the parameters for an equal longest match, preference and metric value. The configuration of multiple static routes, one for each next hop or outbound interface in the case of serial medium is required.

The example demonstrates how two *ip route-static* commands are implemented, each defining the same IP destination address and mask, but alternate next hop locations. This ensures that the longest match (/24) is equal, and naturally so is the preference value, since both routes are static routes that carry a default preference of 60. The cost of both paths is also equal allowing load balancing to occur.

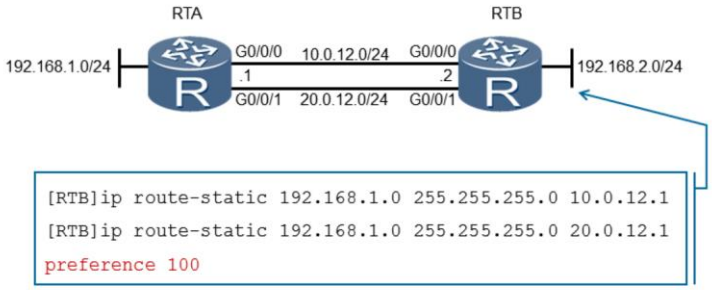
Verifying Static Route Load Balancing

```
[RTB]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

……
192.168.1.0/24      Static  60   0           RD  10.0.12.1       GigabitEthernet 0/0/0
                    Static  60   0           RD  20.0.12.1       GigabitEthernet 0/0/1
```


The routing table can be queried to verify the results by running the *display ip routing-table* command after the static routes are configured. The static route is displayed in the routing table, and results show two entries to the same destination, with matching preference and metric values. The different next hop addresses and variation in the outbound interface identifies the two paths that are taken, and confirms that load balancing has been achieved.

Floating Static Routes



- Floating static routes provide an alternative route in the event that the primary static route fails.

The application of static routes allows for a number of ways that routes can be manipulated to achieve routing requirements. It is possible for the preference of a static route to be changed for the purpose of enabling the preference of one static route over another, or where used with other protocols, to ensure the static route is either preferred or preference is given to the alternative routing protocol.

The default preference value of a static route is 60, therefore by adjusting this preference value, a given static route can be treated with unequal preference over any other route, including other static routes. In the example given, two static routes exist over two physical LAN segments, while normally both static routes would be considered equal, the second route has been given a lesser preference (higher value) causing it to be removed from the routing table. The principle of a floating static route means that the route with a lesser preference will be applied to the routing table, should the primary route ever fail.

Floating Static Route Check

```
[RTB]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
......
192.168.1.0/24  Static  60   0  RD  10.0.12.1 GigabitEthernet0/0/0
```

- Prior to the failure of the primary route, only the primary static route will be present within the routing table.

In using the *display ip routing-table* command, it is possible for the results of the change to the preference value that results in the floating static route, to be observed. Normally two equal cost routes would be displayed in the routing table defining the same destination, however having alternative next hop values and outbound interfaces. In this case however, only one instance can be seen, containing the default static route preference value of 60. Since the second static route now has a preference value of 100, it is not immediately included in the routing table since it is no longer considered an optimal route.

Floating Static Route Check

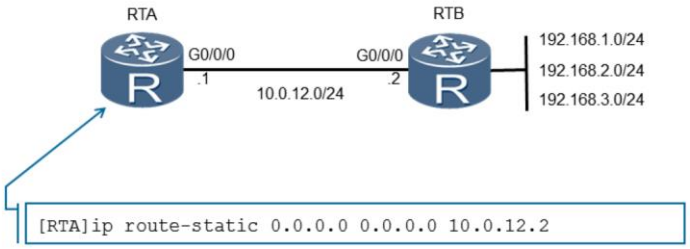
```
[RTB]interface GigabitEthernet 0/0/0
[RTB-GigabitEthernet 0/0/0]shutdown
[RTB]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public  Destinations : 13       Routes : 14
Destination/Mask Proto Pre Cost Flags NextHop Interface
......
192.168.1.0/24 Static  100  0  RD  20.0.12.1 GigabitEthernet 0/0/1
```

- In disabling the primary route, the floating static route is then added to the routing table.

In the event that the primary static route should fail as a result of physical link failure or through the disabling of an interface, the static route will no longer be able to provide a route to the intended destination and therefore will be removed from the routing table. The floating static route is likely to become the next best option for reaching the intended destination, and will be added to the routing table to allow packets to be transmitted over a second alternative path to the intended destination, allowing continuity in light of any failure.

When the physical connection for the original route is restored, the original static route also will take over from the current floating static route, for which the route will be restored in the routing table causing the floating static route to once again await application.

Default Static Routes



- Default routes provide a form of last resort route in the event that no other longest match is found within the routing table.

HUAWEI

The default static route is a special form of static route that is applied to networks in which the destination address is unknown, in order to allow a forwarding path to be made available. This provides an effective means of routing traffic for an unknown destination to a router or gateway that may have knowledge of the forwarding path within an enterprise network.

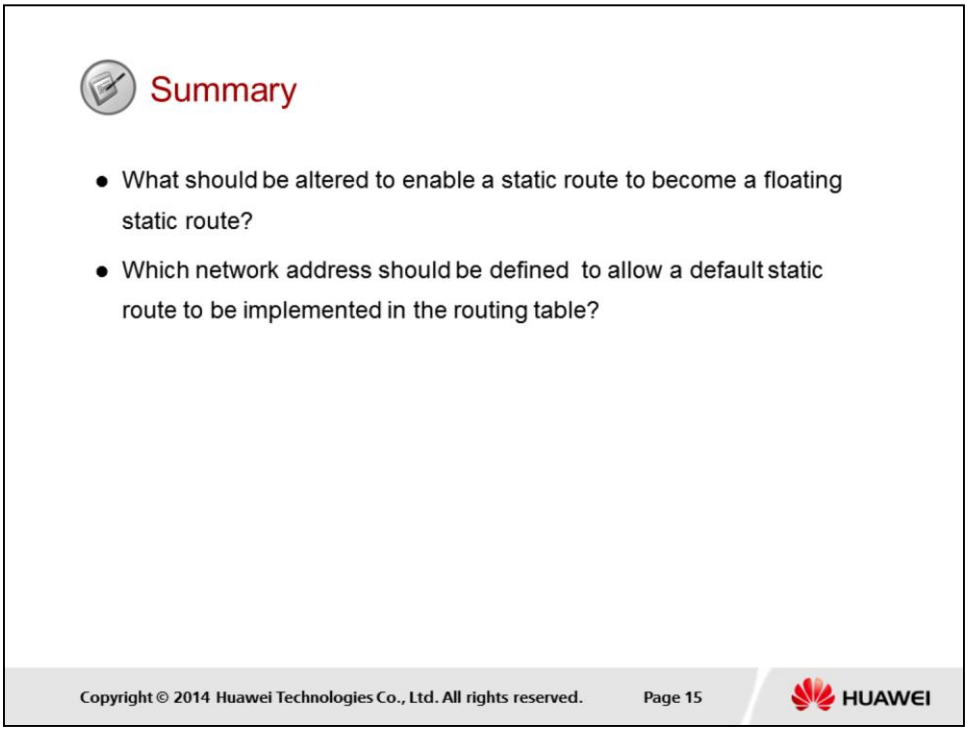
The default route relies on the “any network” address of 0.0.0.0 to match any network to which a match could not be found in the routing table, and provides a default forwarding path to which packets for all unknown network destinations should be routed. In the example, a default static route has been implemented on RTA, identifying that should packets for a network that is unknown be received, such packets should be forwarded to the destination 10.0.12.2.

In terms of routing table decision making, as a static route, the default route maintains a preference of 60 by default, however operates as a last resort in terms of the longest match rule in the route matching process.

Default Static Route Check

```
[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
……
 0.0.0.0/0      Static  60   0  RD   10.0.12.2 GigabitEthernet0/0/0
```


The configuration of the static route once configured will appear within the routing table of the router. The *display ip routing-table* command is used to view this detail. As a result, all routes in the example where not associated with any other routes in the routing table will be forwarded to the next hop destination of 10.0.12.2 via the interface Gigabit Ethernet 0/0/0.



1. A floating static route can be implemented by adjusting the preference value of a static route where two static routes support load balancing.
2. A default static route can be implemented in the routing table by specifying the ‘any network’ address of 0.0.0.0 as the destination address along with a next hop address of the interface to which packets captured by this default static route are to be forwarded.

Thank you

www.huawei.com

Distance Vector Routing with RIP

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

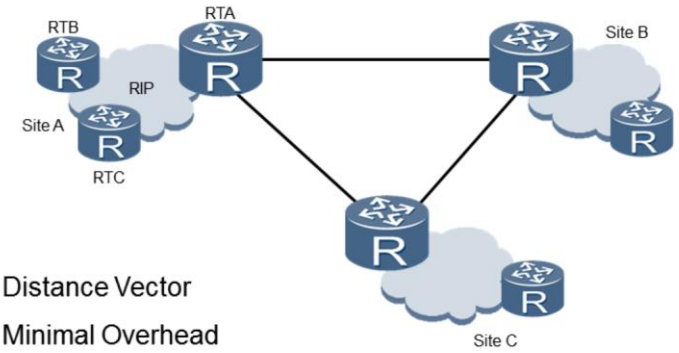
Distance vector routing protocols are a form of dynamic routing protocol that work on the principle of the Bellman-Ford algorithm to define the route that packets should take to reach other network destinations. The application of the Routing Information Protocol (RIP) is often applied in many small networks and therefore remains a valid and popular protocol even though the protocol itself has been in existence much longer than other dynamic routing protocols in use today. The characteristics of such distance vector protocols are represented in this section through the Routing Information Protocol.

Objectives

Upon completion of this section, trainees will be able to:

- Describe the behavior of the Routing Information Protocol.
- Successfully configure RIP routing and associated attributes.

Routing Information Protocol



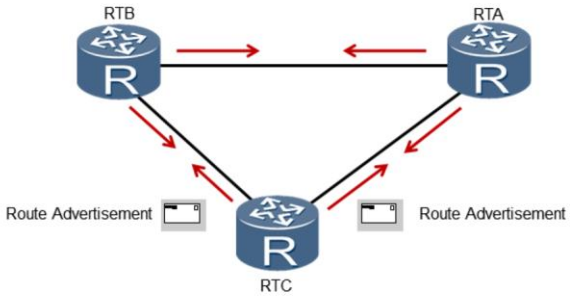
- Distance Vector
- Minimal Overhead
- Suited to Small Networks
- Simple implementation

HUAWEI

The routing information protocol or RIP as it is commonly known, represents one of the more simple forms of routing protocol that are applied to enterprise networks. RIP operates as an interior gateway protocol (IGP) based on the principles of the Bellman-Ford algorithm which operates on the basis of distance vector, defining the path that traffic should take in relation to the optimal distance that is measured using a fixed metric value.

The RIP protocol contains a minimal number of parameters and requires limited bandwidth, configuration and management time, making it ideal for smaller networks. RIP however was not designed with the capability to handle subnets, support interaction with other routing protocols, and provided no means of authentication, since its creation predated the period that these principles were introduced.

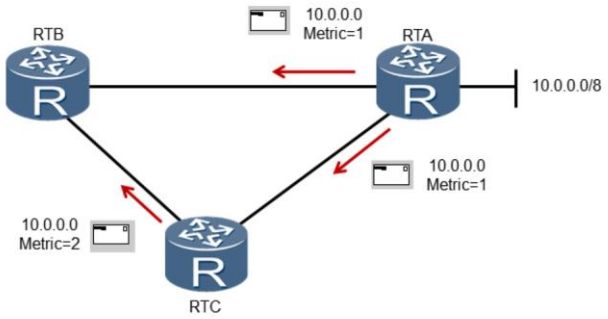
Principle Behavior



- Route Advertisements are sent periodically.
- Advertised information is used to discover the best routes.

Routers that are RIP enabled participate in the advertisement of routing information to neighboring routers. Route advertisements are generated that contain information regarding the networks that are known by the sending router, and the distance to reach those networks. RIP enabled routers advertise to each other, but when they advertise, they only carry the best routing information in their route advertisements.

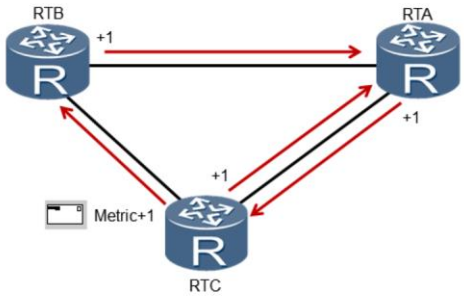
Metrics



- Metric is used to measure the distance to a given network.
- Calculation is based on hops representing a metric of 1.

Each router advertisement contains a number of routes, each associated with a given metric. The metric is used to determine the distance between a router and the destination with which the route advertisement is associated. In RIP the metric is associated with a hop count mechanism where each hop between routers represents a fixed hop count, typically of one. This metric does not take into account any other factors such as the bandwidth for each link or any delay that may be imposed to the link. In the example, router RTB learns of a network via two different interfaces, each providing a hop metric through which, the best route to the destination can be discovered.

Routing Loops and Hop Limits



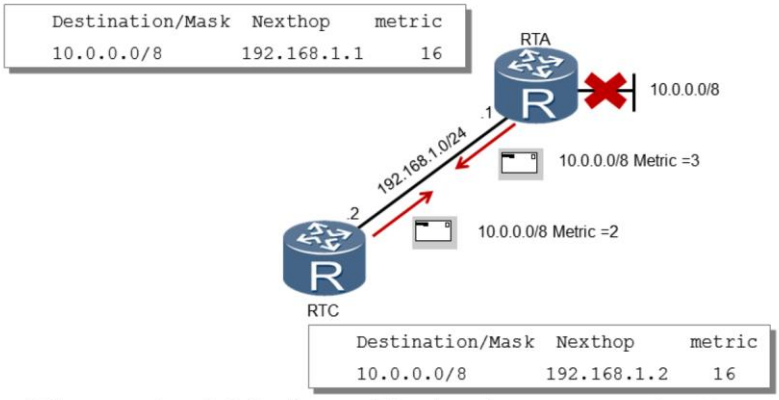
- Metric is incremented by 1 before advertisement is forwarded.
- A limit of 15 hops is defined to prevent infinite forwarding.

As each router processes a route advertisement, the metric value is incremented before forwarding the advertisement to the neighboring router. Where routes become inaccessible however there is a potential for occurrences that results in the hop count becoming infinite.

In order to resolve the problem with infinite route metrics, a value that would represent infinity was defined that allowed the number of possible hops to be restricted to a limit of 15 hops. This metric assumes a network size that is deemed suitable to accommodate the size of networks for which the RIP routing protocol is suited, and also beyond the scale that it is expected any network of this type is expected to reach.

A hop count of 16 would assume the route to be unreachable and cause the network status for the given network to be changed accordingly. Routing loops can occur through a router sending packets to itself, between peering routers or as a result of traffic flow between multiple routers.

Loop Formation

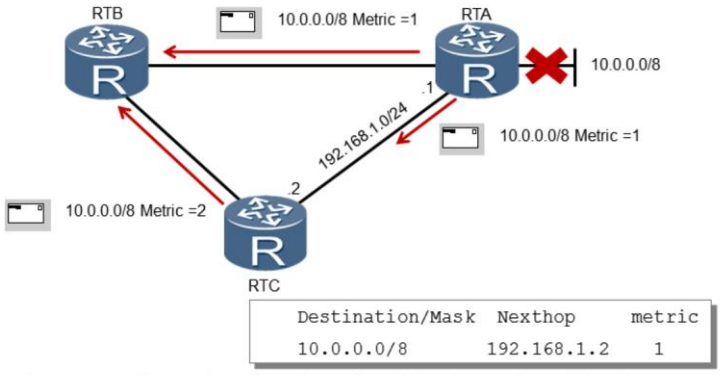


- When a network fails, the next best route may generate a loop.
- A metric of 16 represents an unreachable route.

The example demonstrates how a loop can potentially form where RIP is the routing protocol. A network (10.0.0.0/8) has been learned through the sending of route advertisements from RTA to RTC, for which RTC will have updated its routing table with the network and the metric of 1, in order to reach the destination.

In the event of failure of the connection between router RTA and the network to which it is directly connected, the router will immediately detect loss of the route and consider the route unreachable. Since RTC is possessing knowledge of the network, a route advertisement is forwarded containing information regarding network 10.0.0.0/8. Upon reception of this, RTA will learn of a new route entry for 10.0.0.0/8 with a metric of 2. Since RTC originally learned the route from RTA, any change will need to be updated in RTC also, with a route advertisement being sent to RTC with a metric of 3. This will repeat for an infinite period of time. A metric of 16 allows a cap to be placed on infinity, thereby allowing any route reaching a hop count of 16 to be deemed unreachable.

Loop Prevention-Split Horizon

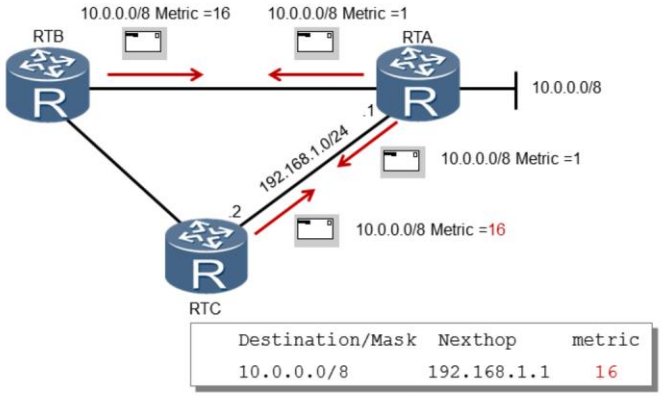


- A route cannot be advertised on the interface via which it was learned.

HUAWEI

Mechanisms have been implemented as part of the RIP routing protocol to address the routing loop issues that occur when routes become inaccessible. One of these mechanisms is known as split horizon and operates on the principle that a route that is learned on an interface, cannot be advertised back over that same interface. This means that network 10.0.0.0/8 advertised to router RTC cannot be advertised back to RTA over the same interface, however will be advertised to neighbors connected via all other interfaces.

Loop Prevention-Poisoned Reverse



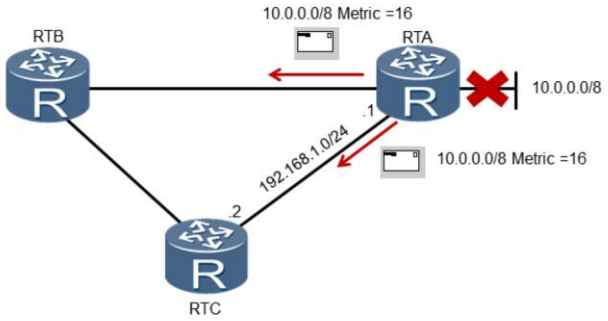
- Poisoned Reverse improves convergence time, however generates additional overhead due to extra route information.

HUAWEI

The implementation of the poison reverse mechanism allows the speed at which erroneous routes are timed out to be increased to almost instantly as a result of allowing routes to be returned to the originating router, containing a metric of 16, to effectively time-out any consideration for a better route where the route becomes invalid.

In the example, RTA advertises a metric of 1 for the network to RTC, while RTC advertises the same network back to RTA to ensure that if 10.0.0.0/8 network fails, RTA will not discover a better path to this network via any other router. This involves however an increase in the size of the RIP routing message, since routes containing the network information received now must also carry the network update, deeming the route unreachable, back to the neighboring router from which the advertisement originated. In Huawei AR2200 series routers, split horizon and poisoned reverse cannot be applied at the same time, if both are configured, only poisoned reverse will be enabled.

Loop Prevention-Triggered Updates

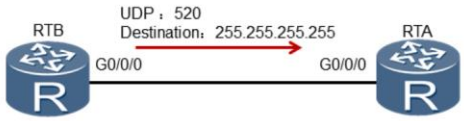


- Updates are sent by default approximately every 30 seconds.
- Triggered updates allow updates to be sent almost instantly.

The default behavior of RIP involves updates of the routing table being sent periodically to neighbors as a route advertisement, which by default is set to occur approximately every 30 seconds. Where links fail however, it also requires that this period be allowed to expire before informing the neighboring routers of the failure.

Triggered updates occur when the local routing information changes and the local router immediately notifies its neighbors of the changes in routing information, by sending the triggered update packet. Triggered updates shorten the network convergence time. When the local routing information changes, the local router immediately notifies its neighbors of the changes in routing information, rather than wait for a periodic update.

RIP Messaging



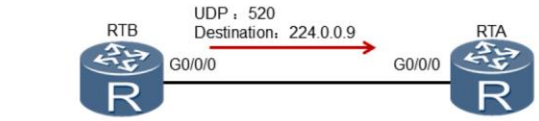
Command	Version	Must be Zero
Address Family Identifier		Must be Zero
IP Address		
Must be Zero		
Must be Zero		
Metric		

RIP is a UDP-based protocol. Each router that uses RIP uses a routing process that involves all communications directed at another router being sent to port 520, including all routing update messages. RIP generally transmits routing update messages as broadcast messages, destined for the broadcast address of 255.255.255.255, referring to all networks. Each router however will generate its own broadcast of routing updates following every update period.

The command and version fields are used once per packet, with the command field detailing whether the packet is a request or response message, for which all update messages are considered response messages. The version refers to the version of RIP, which in this case is version 1. The remaining fields are used to support the network advertisements for which up to 25 route entries can be advertised in a single RIP update message.

The address family identifier lists the protocol type that is being supported by RIP, which in this example is IP. The remaining fields are used to carry the IP network address and the hop metric that contains a value between 1 and 15 (inclusive) and specifies the current metric for the destination; or the value 16 (infinity), which indicates that the destination is not reachable.

RIP Extensions



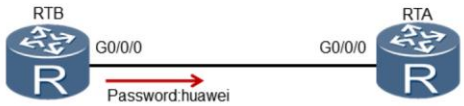
Command	Version	Unused
Address Family Identifier		Route Tag
IP Address		
Subnet Mask		
Next Hop		
Metric		

The introduction of a new version of RIP, known as RIP version 2, does not change RIP as such but rather provides extensions to the current RIP protocol to allow for a number of ambiguities to be resolved. The format of the RIP datagram applies the same principles of the original RIP protocol with the same command parameters. The version field highlights the extended fields are part of version 2.

The address family identifier continues to refer to the protocol being supported and also may be used in support of authentication information as explained shortly. The route tag is another feature that is introduced to resolve limitations that exist with support for interaction between autonomous systems in RIP, the details of which however fall outside of the scope of this course. Additional parameter extensions have been made part of the route entry including the Subnet Mask field which contains the subnet mask that is applied to the IP address, to define the network or sub-network portion of the address.

The Next Hop field now allows for the immediate next hop IP address, to which packets destined for the destination address specified in a route entry, should be forwarded. In order to reduce the unnecessary load of hosts that are not listening for RIP version 2 packets, an IP multicast address is used to facilitate periodic broadcasts, for which the IP multicast address used is 224.0.0.9.

RIP Extensions – Authentication



Command	Version	Unused
0XFFFF		Authentication Type
Authentication		

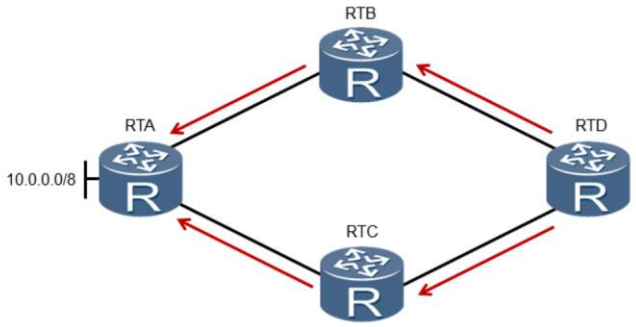
- RIP version 2 allows for authentication between peers.
- Supports plaintext and cryptographic authentication.

Authentication represents a means by which malicious packets can be filtered, by ensuring that all packets received can be verified as originating from a valid peer through the use of a key value. This key value originally represents a plaintext password string that can be configured for each interface, as recognized by the authentication type of 2. The authentication configured between peers must match before RIP messages can be successfully processed. For authentication processing, if the router is not configured to authenticate RIP version 2 messages, then RIP version 1 and unauthenticated RIP version 2 messages will be accepted; authenticated RIP version 2 messages shall be discarded.

If the router is configured to authenticate RIP version 2 messages, then RIP version 1 messages and RIP version 2 messages which pass authentication testing shall be accepted; unauthenticated and failed authentication RIP version 2 messages shall be discarded.

RIP version 2 originally supported only simple plaintext authentication that provided only minimal security since the authentication string could easily be captured. With the increased need for security for RIP, cryptographic authentication was introduced, initially with the support for a keyed-MD5 authentication (RFC 2082) and further enhancement through the support of HMAC-SHA-1 authentication, introduced as of RFC 4822. While Huawei AR2200 series routers are capable of supporting all forms of authentication mentioned, the example given demonstrates the original authentication for simplicity.

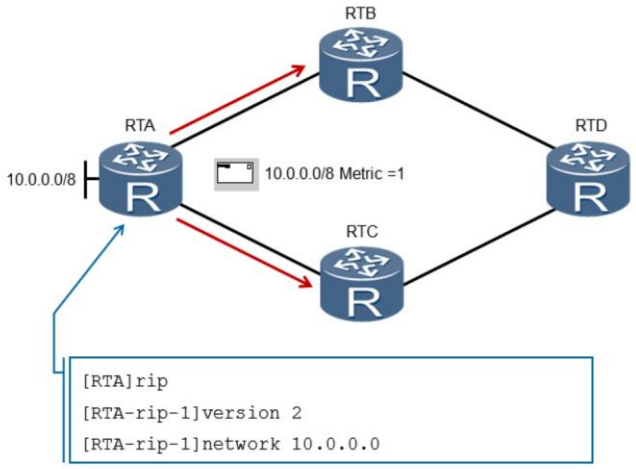
RIP Load Balancing



- Load balancing can be used in RIP to utilize redundant links.
- AR2200 supports up to 8 equal cost routes by default.

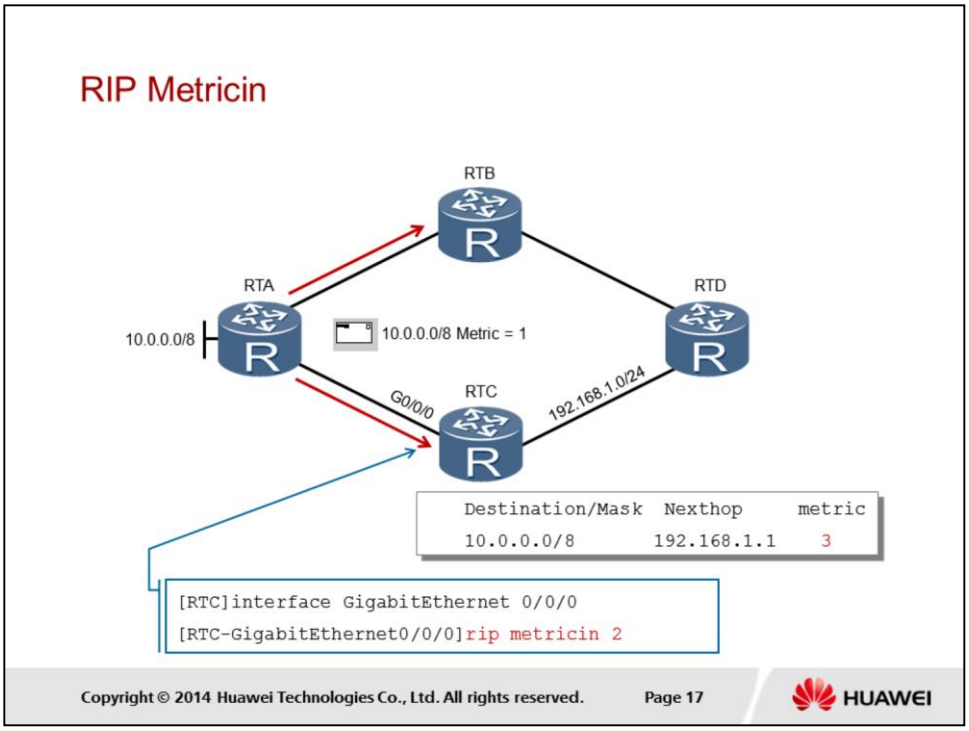
If a network has multiple redundant links, a maximum number of equal-cost routes can be configured to implement load balancing. In this manner, network resources are more fully utilized, situations where some links are overloaded while others are idle can be avoided, and long delays in packet transmissions can be prevented. The default and maximum number of equal cost routes supported by RIP is 8 at any one time.

RIP Network Advertisement



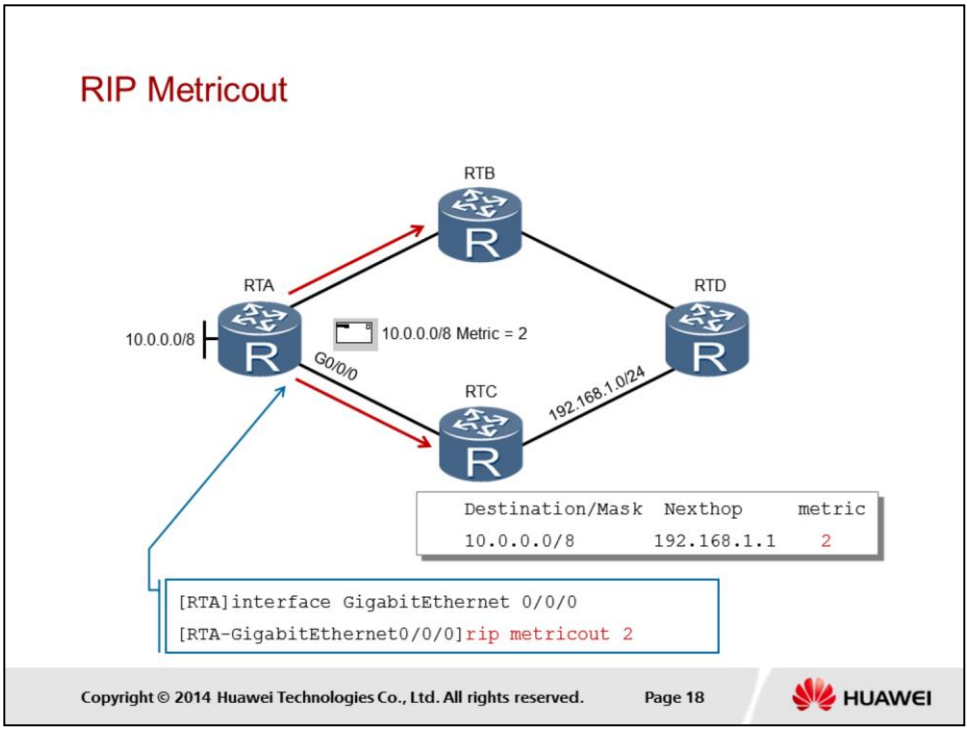
It is required for all routers supporting the RIP routing process to first enable the process on each router. The *rip* [*process-id*] command is used to enable this, with the process-id identifying a specific process ID to which the router is associated. If the process ID is not configured, the process will default to a process ID of 1.Where variation in the process ID exists, the local router will create separate RIP routing table entries for each process that is defined.

The *version 2* command enables the RIP version 2 extension to RIP allowing for additional capability for subnets, authentication, inter-autonomous system communication etc. The *network <network-address>* command specifies the network address for which RIP is enabled, and must be the address of the natural network segment.



RIP is also capable of supporting manipulation of RIP metrics to control the flow of traffic within a RIP routing domain. One means to achieve this is to adjust the metric associated with the route entry when received by a router. When an interface receives a route, RIP adds the additional metric of the interface to the route, and then installs the route into the routing table, thereby increasing the metric of an interface which also increases the metric of the RIP route received by the interface.

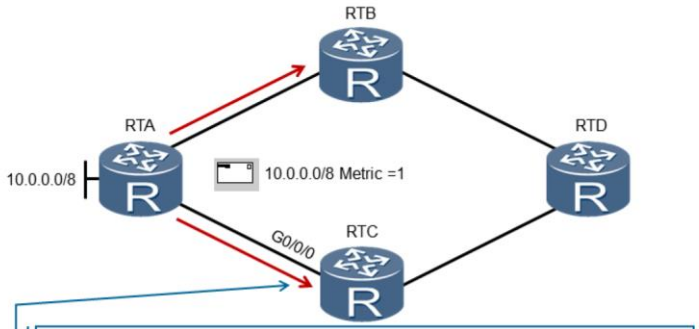
The rip metricin <metric value> command allows for manipulation of the metric, where the metric value refers to the metric that is to be applied. It should also be noted that for the rip metricin command the metric value is added to the metric value currently associated with the route. In the example, the route entry for network 10.0.0.0/8 contains a metric of 1, and is manipulated upon arrival at the interface of RTC, resulting in the metric value of 3 being associated with the route.



The *rip metricout* command allows for the metric to be manipulated for the route when a RIP route is advertised. Increasing the metric of an interface also increases the metric of the RIP route sent on the interface but does not affect the metric of the route in the routing table of the router to which the *rip metricout* command is applied.

In its most basic form the *rip metricout* command defines the value that must be adopted by the forwarded route entry, but is also capable of supporting filtering mechanisms to selectively determine to which routes the metric should be applied. The general behavior of RIP is to increment the metric by one before forwarding the route entry to the next hop. If the *rip metricout* command is configured, only the metric value referenced in the command is applied.

Split Horizon & Poisoned Reverse




```
[RTC]interface GigabitEthernet 0/0/0
[RTC-GigabitEthernet0/0/0]rip split-horizon
[RTC-GigabitEthernet0/0/0]rip poison-reverse
```

- If both are enabled, only *rip poison-reverse* will take effect.

The configuration of both split horizon and poisoned reverse are performed on a per interface basis, with the *rip split-horizon* command being enabled by default (with exception to NBMA networks) in order to avoid many of the routing loop issues that have been covered within this section. The implementation of both split horizon and poisoned reverse is not permitted on the AR2200 series router, therefore where poisoned reverse is configured on the interface using the *rip poison-reverse* command, split horizon will be disabled.

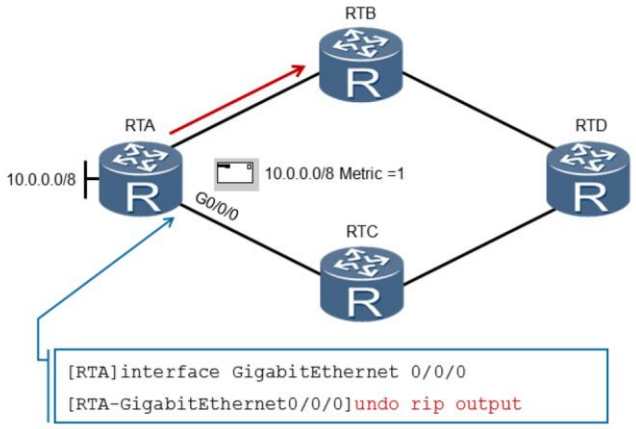
Configuration Validation

```
[RTC] display rip 1 interface GigabitEthernet0/0/0 verbose
 GigabitEthernet0/0/0(192.168.1.2)
  State           : UP          MTU    : 500
  Metricin        : 2
  Metricout       : 1
  Input           : Enabled     Output : Enabled
  Protocol        : RIPv2 Multicast
  Send version    : RIPv2 Multicast Packets
  Receive version : RIPv2 Multicast and Broadcast Packets
  Poison-reverse                : Enabled
  Split-Horizon                 : Enabled
  Authentication type           : None
  Replay Protection             : Disabled
```

- Both show as enabled but only "*Poison-reverse*" will take effect.

The configuration of the routing information protocol on a per interface basis can be verified through the *display rip <process_id> interface <interface> verbose* command. The associated RIP parameters can be found in the displayed output, including the RIP version applied along with other parameters such as whether poison-reverse and split-horizon have been applied to the interface. Where the display command references that both the poison-reverse and split-horizon as both enabled, only the poison-reverse command will take effect.

RIP Output

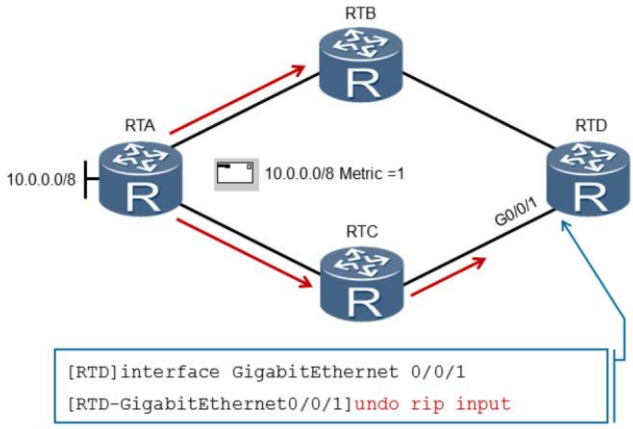


- Outbound RIP advertisements restricted on the G0/0/0 interface.

HUAWEI

The *rip output* command is applied to the interface of a router participating in RIP routing and allows RIP to forward update messages out from the interface. Where the *undo rip output* command is applied to an interface, the RIP update message will cease to be forwarded out from a given interface. It's application is valid in circumstances where an enterprise network wishes to not share its internal routes via an interface that connects to an external network in order to protect the network, often applying a default route to this interface instead for any routes which wish to reach external networks.

RIP Input



- Inbound RIP advertisements restricted on the G0/0/1 interface.

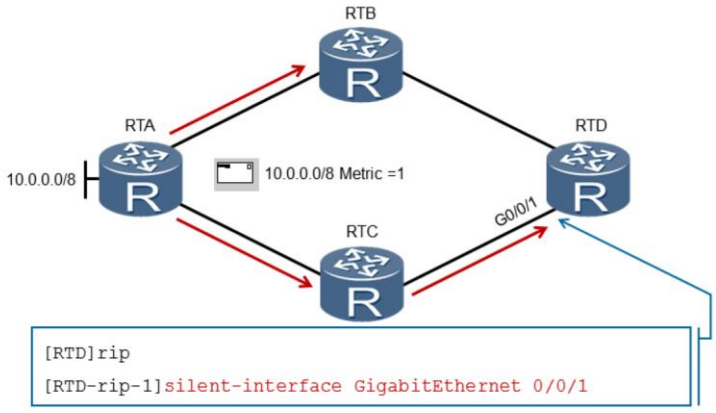
The *undo rip input* command allows an interface to reject all RIP update messages and prevent RIP information from being added to the routing table for a given interface. This may be applied in situations where the flow of traffic may require to be controlled via certain interfaces only, or prevent RIP from being received by the router completely. As such any RIP update messages sent to the interface will be discarded immediately. The *rip input* command can be used to re-enable an interface to resume receipt of RIP updates.

Configuration Validation

```
[RTD] display rip 1 interface GigabitEthernet0/0/1 verbose
 GigabitEthernet0/0/1(192.168.1.2)
  State           : UP          MTU    : 500
  Metricin        : 1
  Metricout       : 1
  Input           : Disabled     Output : Enabled
  Protocol        : RIPv2 Multicast
  Send version    : RIPv2 Multicast Packets
  Receive version : RIPv2 Multicast and Broadcast Packets
  Poison-reverse                : Enabled
  Split-Horizon                 : Enabled
  Authentication type           : None
  Replay Protection             : Disabled
```


The *display rip <process_id> interface <interface> verbose* command can also be used to confirm the implementation of restrictions to the interface. Where the interface has been configured with the undo rip input, the capability to receive RIP routes will be considered disabled as highlighted under the Input parameter.

Silent Interface



- Interface will not participate in RIP, but will receive RIP routes.
- Takes precedence over *rip input* and *rip output* commands

HUAWEI

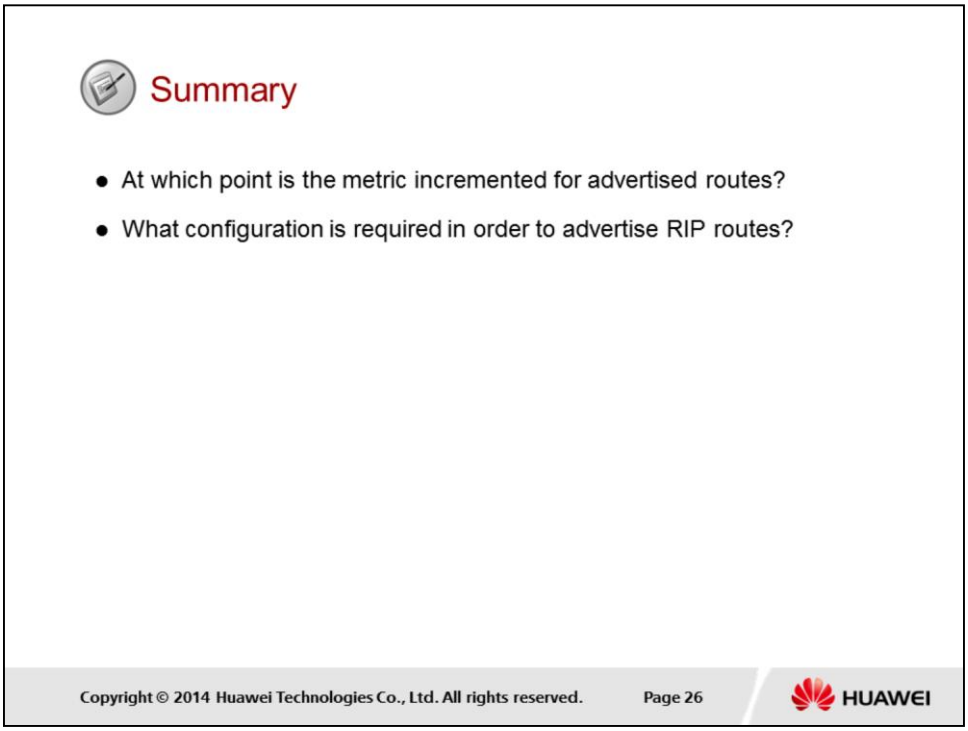
The silent interface allows for RIP route updates to be received and used to update the routing table of the router, but will not allow an interface to participate in RIP. In comparison, the silent-interface command has a higher precedence than both rip input & rip output commands. Where the silent-interface all command is applied, the command takes the highest priority, meaning no single interface can be activated. The silent-interface command must be applied per interface to allow for a combination of active and silent interfaces.

A common application of the silent interface is for non broadcast multi access networks. Routers may be required to receive RIP update messages but wish not to broadcast/multicast its own updates, requiring instead that a relationship with the peering router be made through the use of the peer <ip address> command.

Configuration Validation

```
[RTD] display rip
Public VPN-instance
    RIP process : 1
       RIP version    : 2
       Preference     : 100
       Checkzero      : Enabled
       Default-cost   : 0
       Summary        : Enabled
       Host-route     : Enabled
       Maximum number of balanced paths : 8
       Update time    : 30 sec              Age time : 180 sec
       Garbage-collect time : 120 sec
       Graceful restart  : Disabled
       BFD               : Disabled
       Silent-interfaces : GigabitEthernet0/0/1
```


The display rip command provides a more comprehensive router based output for which global parameters can be verified along with certain interface based parameters. The implementation of the silent-interface command on a given interface for example can be observed through this command.



1. The metric is incremented prior to the forwarding of the route advertisement from the outbound interface.
2. The advertisement of RIP routes is achieved through the configuration of the network command. For each network that is to be advertised by a router, a network command should be configured.

Thank you

www.huawei.com

Link State Routing with OSPF

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

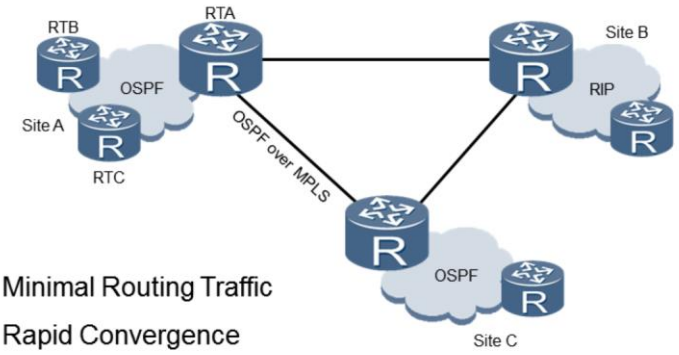
OSPF is an interior gateway protocol (IGP) designed for IP networks, that is founded on the principles of link state routing. The link state behavior provides many alternative advantages for medium and even large enterprise networks. Its application as an IGP is introduced along with information relevant to the understanding of OSPF convergence and implementation, for supporting OSPF in enterprise networks.

Objectives

Upon completion of this section, trainees will be able to:

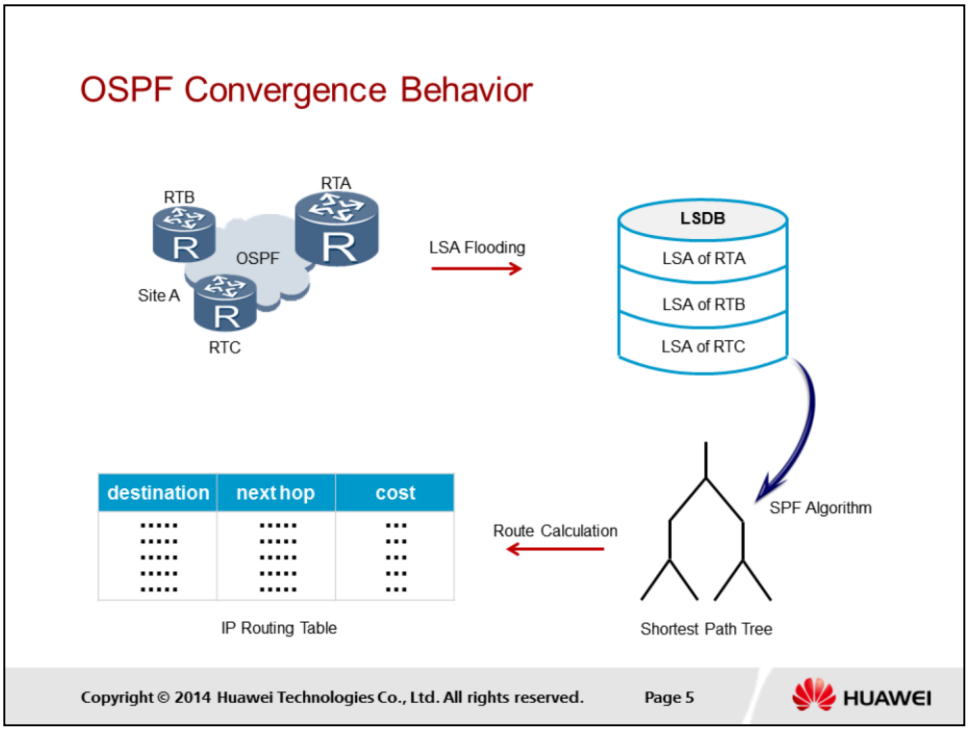
- Explain the OSPF convergence process
- Describe the different network types supported by OSPF
- Successfully configure single area OSPF networks

Open Shortest Path First（OSPF）



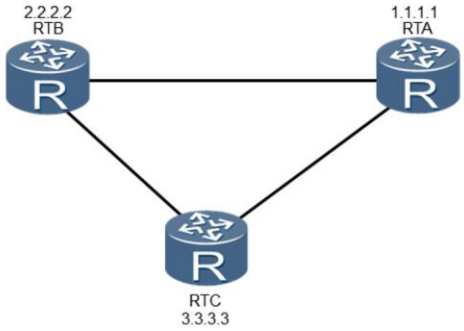
- Minimal Routing Traffic
- Rapid Convergence
- Scalable
- Accurate Route Metrics

Open Shortest Path First or OSPF is regarded as a link state protocol that is capable of quickly detecting topological changes within the autonomous system and establish loop free routes in a short period of time, with minimum additional communication overhead for negotiating topology changes between peering routers. OSPF also deals with scalability issues that occur when communication between an expanding number of routers becomes so extreme that it begins to lead to instability within the autonomous system. This is managed through the use of areas that limits the scope of router communication to an isolated group within the autonomous system allowing small, medium and even large networks to be supported by OSPF. The protocol is also able to work over other protocols such as MPLS, a label switching protocol, to provide network scalability even over geographically disperse locations. In terms of optimal path discovery, OSPF provides rich route metrics that provides more accuracy than route metrics applied to protocols such as RIP to ensure that routes are optimized, based on not only distance but also link speed.



The convergence of OSPF requires that each and every router actively running the OSPF protocol have knowledge of the state of all interfaces and adjacencies (relationship between the routers that they are connected to), in order to establish the best path to every network. This is initially formed through the flooding of Link State Advertisements (LSA) which are units of data that contain information such as known networks and link states for each interface within a routing domain. Each router will use the LSA received to build a link state database (LSDB) that provides the foundation for establishing the shortest path tree to each network, the routes from which are ultimately incorporated into the IP routing table.

Router ID

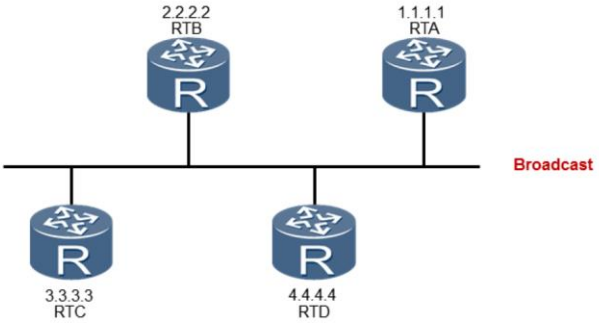


- A router ID is a 32 bit value used to identify each router running the OSPF protocol.

The router ID is a 32-bit value assigned to each router running the OSPF protocol. This value uniquely identifies the router within an Autonomous System. The router ID can be assigned manually, or it can be taken from a configured address. If a logical (loopback) interface has been configured, the router ID will be based upon the IP address of the highest configured logical interface, should multiple logical interfaces exist.

If no logical interfaces have been configured, the router will use the highest IP address configured on a physical interface. Any router running OSPF can be restarted using the graceful restart feature to renew the router ID should a new router ID be configured. It is recommended that the router ID be configured manually to avoid unexpected changes to the router ID in the event of interface address changes.

OSPF Supported Network Types

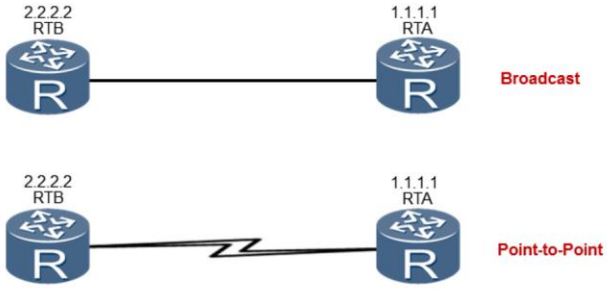


- Ethernet based networks adopt the broadcast network type by default.

HUAWEI

OSPF supports various network types, and in each case will apply a different behavior in terms of how neighbor relationships are formed and how communication is facilitated. Ethernet represents a form of broadcast network that involves multiple routers connected to the same network segment. One of the primary issues faced regards how communication occurs between the neighboring routers in order to minimize OSPF routing overhead. If an Ethernet network is established, the broadcast network type will be applied automatically in OSPF.

OSPF Supported Network Types

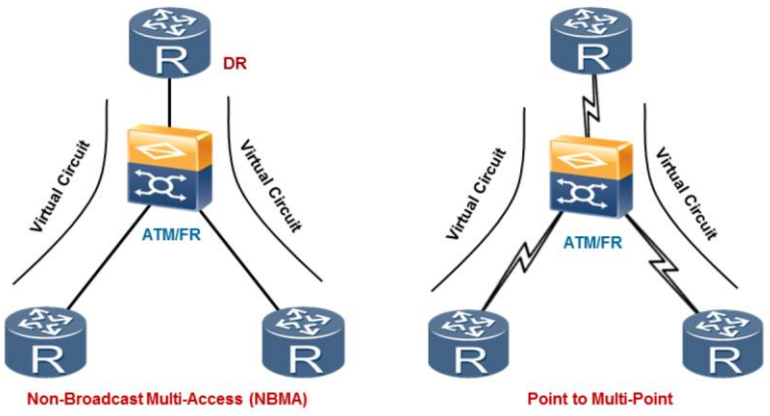


- Serial technologies such as PPP and HDLC will default to the Point-to-Point network type.

HUAWEI

Where two routers are established in a point-to-point topology, the applied network type will vary depending on the medium and link layer technology applied. As mentioned, the use of an Ethernet medium will result in the broadcast network type for OSPF being assigned automatically. Where the physical medium is serial, the network type is considered point-to-point. Common forms of protocols that operate over serial media at the link layer include Point to Point Protocol (PPP) and High-level Data Link Control (HDLC).

OSPF Supported Network Types



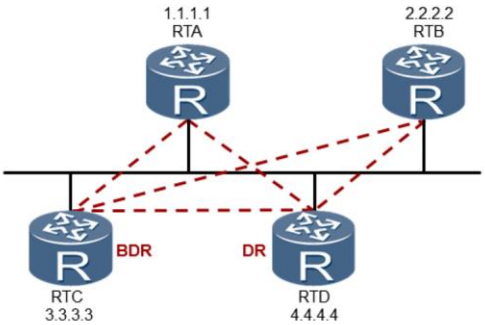
- ATM & Frame Relay default to Non-Broadcast Multi-Access.

HUAWEI

OSPF may operate over multi access networks that do not support broadcasts. Such networks include Frame Relay and ATM that commonly operate using hub and spoke type topologies, which rely on the use of virtual circuits in order for communication be to achieved. OSPF may specify two types of networks that can be applied to links connected to such environments. The Non Broadcast Multi Access (NBMA) network type emulates a broadcast network and therefore requires each peering interface be part of the same network segment. Unlike a broadcast network, the NBMA forwards OSPF packets as a unicast, thereby requiring multiple instances of the same packet be generated for each destination.

Point-to-Multipoint may also be applied as the network type for each interface, in which case a point-to-point type behavior is applied. This means that each peering must be associated with different network segments. Designated Routers are associated with broadcast networks, and therefore are implemented by NBMA networks. Most importantly is the positioning of a DR which must be assigned on the hub node of the hub and spoke architecture to ensure all nodes can communicate with the DR.

Designated Router & Backup Designated Router

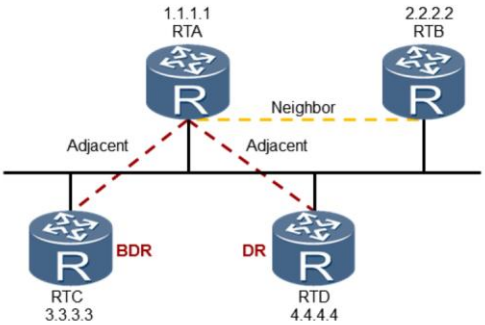


- Designated Routers limit the number of adjacencies necessary in broadcast (Ethernet) networks.

In order to address and optimize the communication of OSPF over broadcast networks, OSPF implements a Designated Router (DR) that acts as a central point of communication for all other routers associated with a broadcast network on at least one interface. In a theoretical broadcast network that does not apply a DR, it can be understood that the communication follows an n(n-1)/2 formula, where n represents the number of router interfaces participating in OSPF. In the example given, this would refer to 6 adjacencies between all routers. When the DR is applied, all routers establish a relationship with the DR to which is responsible for acting as a central point of communication for all neighboring routers in a broadcast network.

A Backup Designated Router (BDR) is a router that is elected to take over from the DR should it fail. As such it is necessary that the BDR establish a link state database as that of the DR to ensure synchronization. This means that all neighboring routers must also communicate with the BDR in a broadcast network. With the application of the DR and BDR, the number of associations is reduced from 6 to 5 since RTA and RTB need only communicate with the DR and BDR. This may appear to have a minimal effect however where this is applied to a network containing for example 10 routers, i.e. (10*9)/2 the resulting communication efficiency becomes apparent.

Neighbor States

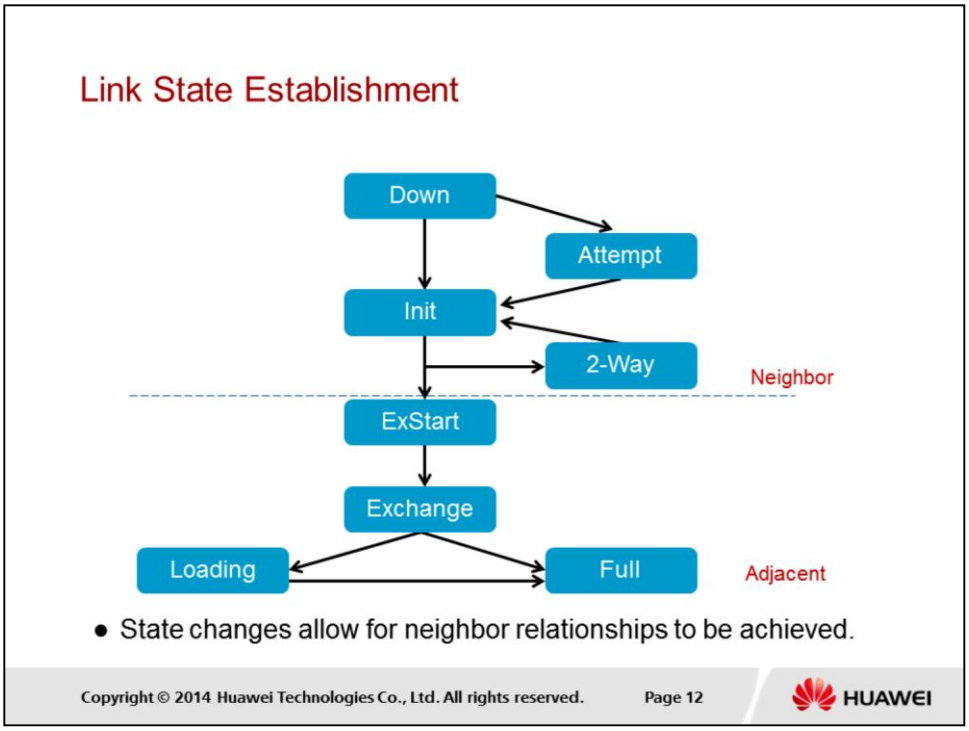


- Defines form of relationship between neighbors.
- Two neighbor states are possible, neighbor and adjacent.

HUAWEI

OSPF creates adjacencies between neighboring routers for the purpose of exchanging routing information. Not every two neighboring routers will become adjacent, particularly where one of the two routers establishing an adjacency is considered to not be the DR or BDR. These routers are known as DROther and only acknowledge the presence of the DROther but do not establish full communication; this state is known as the neighbor state. DROther routers do however form full adjacency with both DR and BDR routers to allow synchronization of the link state database of the DR and BDR routers with each of the DROther routers. This synchronization is achieved by establishing an adjacent state with each DROther.

An adjacency is bound to the network that the two routers have in common. If two routers have multiple networks in common, they may have multiple adjacencies between them.

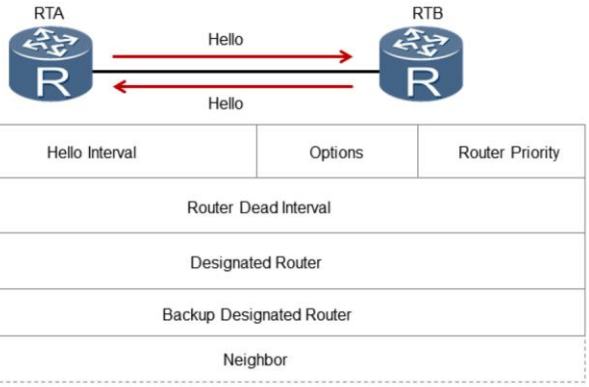


Each router participating in OSPF will transition through a number of link states to achieve either a neighbor state or an adjacent state. All routers begin in the down state upon initialization and go through a neighbor discovery process, which involves firstly making a routers presence known within the OSPF network via a Hello packet. In performing this action the router will transition to an init state.

Once the router receives a response in the form of a Hello packet containing the router ID of the router receiving the response, a 2-way state will be achieved and a neighbor relationship formed. In the case of NBMA networks, an attempt state is achieved when communication with the neighbor has become inactive and an attempt is being made to re-establish communication through periodic sending of Hello packets. Routers that do not achieve an adjacent relationship will remain in a neighbor state with a 2-way state of communication.

Routers such as DR and BDR will build an adjacent neighbor state with all other neighboring routers, and therefore must exchange link state information in order to establish a complete link state database. This requires that peering routers that establish an adjacency first negotiate for exchange of link state information (ExStart) before proceeding to exchange summary information regarding the networks they are aware of. Neighbors may identify routes they are either not aware of or do not have up to date information for, and therefore request additional details for these routes as part of the loading state. A fully synchronized relationship between neighbors is determined by the full state at which time both peering routers can be considered adjacent.

Neighbor Discovery



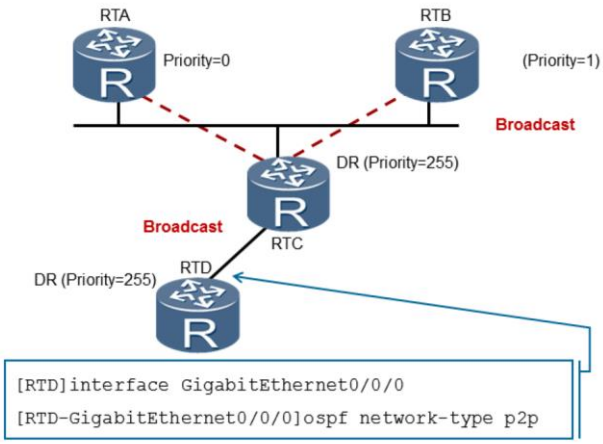
- The Hello protocol is responsible for neighbor discovery and maintenance for two way communication between neighbors.

HUAWEI

Neighbor discovery is achieved through the use of Hello packets that are generated at intervals based on a Hello timer, which by default is every 10 seconds for broadcast and point to point network types; whereas for NBMA and Point to Multipoint network types the hello interval is 30 seconds. The hello packet contains this interval period, along with a router priority field that allows neighbors to determine the neighbor with the highest router ID for identification of the DR and BDR in broadcast and NBMA networks.

A period specifying how long a hello packet is valid before the neighbor is considered lost must also be defined, and this is carried as the router dead interval within the hello packet. This dead interval is set by default to be four times the hello interval, thus being 40 seconds for broadcast and point to point networks, and 120 seconds for NBMA and Point to Multipoint networks. Additionally, the router ID of both the DR and BDR are carried, where applicable, based on the network for which the hello packet is generated.

Designated Router Election



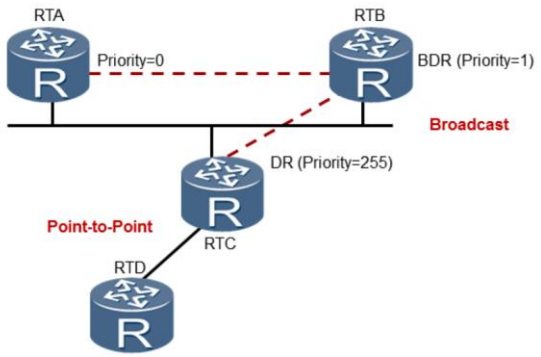
- A Designated Router is elected based on the priority value.

HUAWEI

Following neighbor discovery, the DR election may occur depending on the network type of the network segment. Broadcast and NMBA networks will perform DR election. The election of the DR relies on a priority that is assigned for each interface that participates in the DR election process. This priority value is set as 1 by default and a higher priority represents a better DR candidate.

If a priority of 0 is set, the router interface will no longer participate in the election to become the DR or BDR. It may be that where point to point connections (using Ethernet as the physical medium) are set to support a broadcast network type, unnecessary DR election will occur, which generates excessive protocol traffic. It therefore is recommended that the network type be configured as a point to point network type.

Backup Designated Router Election



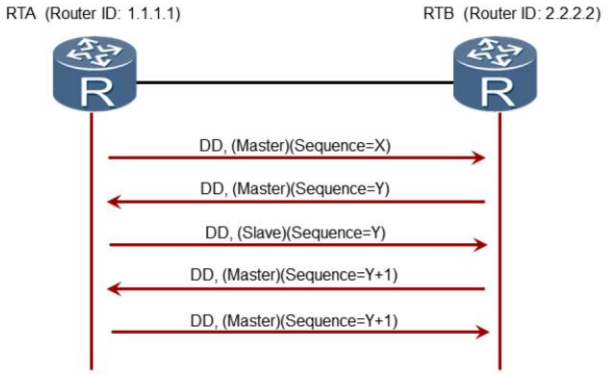
- The Backup Designated Router (BDR) forms adjacencies with all other routers and will become the DR if the existing DR fails.

HUAWEI

In order to make the improve the efficiency of transition to a new Designated Router, a Backup Designated Router is assigned for each broadcast and NBMA network. The Backup Designated Router is also adjacent to all routers on the network, and becomes the Designated Router when the previous Designated Router fails. If there were no Backup Designated Router present, new adjacencies would have to be formed between the new Designated Router and all other routers attached to the network.

Part of the adjacency forming process involves the synchronizing of link-state databases, which can potentially take quite a long time. During this time, the network would not be available for the transit of data. The Backup Designated Router obviates the need to form these adjacencies, since they already exist. This means the period of disruption in transit traffic lasts only as long as it takes to flood the new LSAs (which announce the new Designated Router). The Backup Designated Router is also elected by the Hello packet. Each Hello packet has a field that specifies the Backup Designated Router for the network.

Database Synchronization

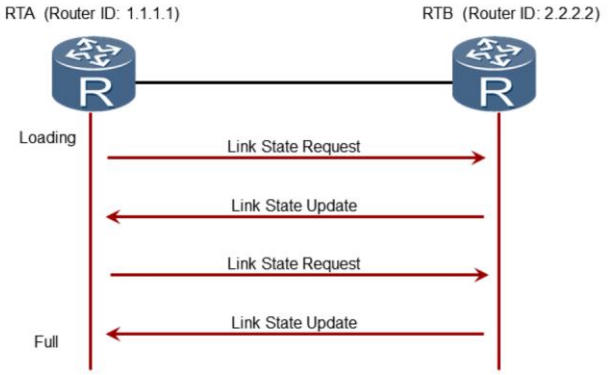


- Neighboring routers form a master/slave relationship.
- Database Description packets contain LSA header information.

In a link-state routing algorithm, it is very important for all routers' link-state databases to stay synchronized. OSPF simplifies this by requiring only adjacent routers remain synchronized. The synchronization process begins as soon as the routers attempt to bring up the adjacency. Each router describes its database by sending a sequence of Database Description packets to its neighbor. Each Database Description packet describes a set of LSAs belonging to the router's database.

When the neighbor sees an LSA that is more recent than its own database copy, it makes a note that this newer LSA should be requested. This sending and receiving of Database Description packets is called the "Database Exchange Process". During this process, the two routers form a master/slave relationship. Each Database Description packet has a sequence number. Database Description packets sent by the master are acknowledged by the slave through echoing of the sequence number.

Establishing Full Adjacency

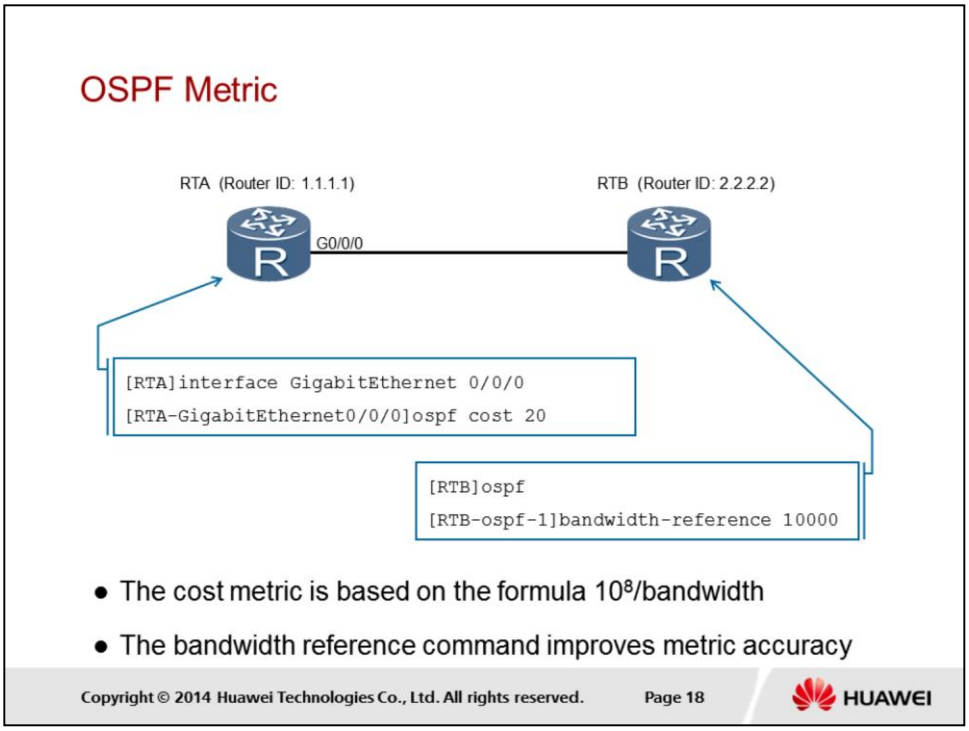


- Missing or newer instances of LSA are requested using LSR
- The entire requested LSA is sent as an update.

During and after the Database Exchange Process, each router has a list of those LSAs for which the neighbor has more up-to-date instances. The Link State Request packet is used to request the pieces of the neighbor's database that are more up-to-date. Multiple Link State Request packets may need to be used.

Link State Update packets implement the flooding of LSAs. Each Link State Update packet carries a collection of LSAs one hop further from their origin. Several LSAs may be included in a single packet. On broadcast networks, the Link State Update packets are multicast. The destination IP address specified for the Link State Update Packet depends on the state of the interface. If the interface state is DR or Backup, the address AllSPFRouters (224.0.0.5) should be used. Otherwise, the address AllDRouters (224.0.0.6) should be used. On non-broadcast networks, separate Link State Update packets must be sent, as unicast, to each adjacent neighbor (i.e. those in a state of Exchange or greater). The destination IP addresses for these packets are the neighbors' IP addresses.

When the Database Description Process has completed and all Link State Requests have been satisfied, the databases are deemed synchronized and the routers are marked fully adjacent. At this time the adjacency is fully functional and is advertised in the two routers' router-LSAs.

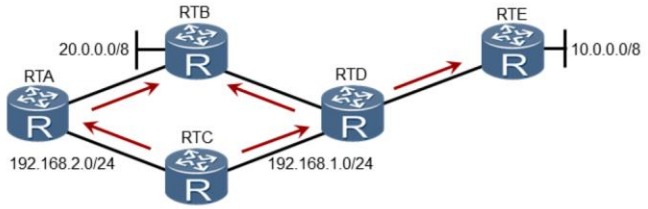


OSPF calculates the cost of an interface based on bandwidth of the interface. The calculation formula is: cost of the interface=reference value of bandwidth/bandwidth. The reference value of bandwidth is configurable for which the default is 100 Mbps. With the formula 100000000/Bandwidth, this gives a cost metric of 1562 for a 64 kbit/s Serial port, 48 for an E1 (2.048 Mbit/s) interface and a cost of 1 for Ethernet (100 Mbit/s) or higher.

To be able to distinguish between higher speed interfaces it is imperative that the cost metric be adjusted to match the speeds currently supported. The bandwidth-reference commands allows the metric to be altered by changing the reference value of the bandwidth in the cost formula. The higher the value, the more accurate the metric. Where speeds of 10Gb are being supported, it is recommended that the bandwidth-reference value be increased to '10000' or 10^{10}/bandwidth to provide metrics of 1, 10 and 100 for 10Gb, 1Gb and 100Mb bandwidth links respectively.

Alternatively the cost can be manually configured by using the ospf cost command to define a cost value for a given interface. The cost value ranges from 1 to 65535 with a default cost value of 1.

Shortest Path Tree



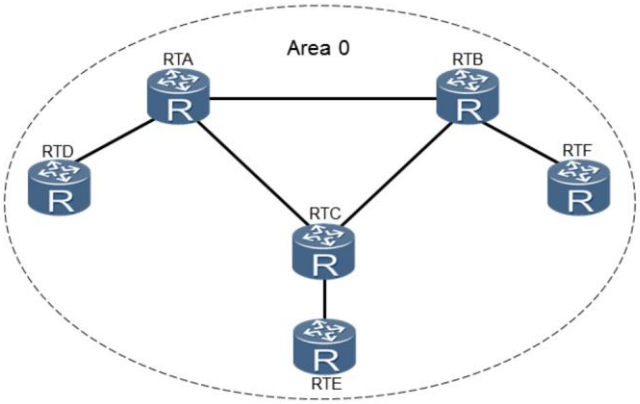

```
[RTC]display ip routing-table
......
Destination/Mask Proto Pre Cost   Flags NextHop       Interface
 10.0.0.0/8      OSPF  10  20       D   192.168.1.4    G0/0/0
 20.0.0.0/8      OSPF  10  20       D   192.168.1.4    G0/0/0
                 OSPF  10  20       D   192.168.2.1    G0/0/1
```

- Each router calculates the shortest path to all other networks

HUAWEI

A router that has achieved a full state is considered to have received all link state advertisements (LSA) and synchronized its link state database (LSDB) with that of the adjacent neighbors. The link state information collected in the link state database is then used to calculate the shortest path to each network. Each router only relies on the information in the LSDB in order to independently calculate the shortest path to each destination, as opposed to relying on select route information from peers which is deemed to be the best route to a destination. The calculation of the shortest path tree however means that each router must utilize additional resources to achieve this operation.

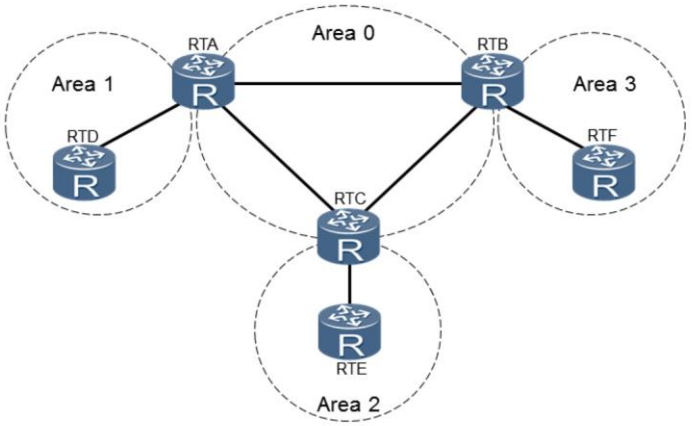
OSPF Areas – Single Area



- A single link state database for the administrative domain.
- Any area number can be assigned but area 0 is recommended.

Smaller networks may involve a select number of routers which operate as part of the OSPF domain. These routers are considered to be part of an area which is represented by an identical link state database for all routers within the domain. As a single area, OSPF can be assigned any area number, however for the sake of future design implementation it is recommended that this area be assigned as area 0.

OSPF Areas – Multi Area

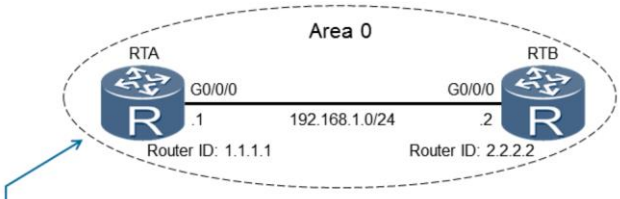


- Areas build separate LS databases, minimize impact of change

The need to forward link state advertisements and subsequent calculation of the shortest path based on the link state database becomes increasingly complex as more and more routers become a part of the OSPF domain. As such, OSPF is capable of supporting a hierarchical structure to limit the size of the link state database, and the number of calculations that must be performed when determining the shortest path to a given network.

The implementation of multiple areas allows an OSPF domain to compartmentalize the calculation process based on a link state database, that is only identical for each area, but provides the information to reach all destinations within the OSPF domain. Certain routers known as area border routers (ABR) operate between areas and contain multiple link state databases for each area that the ABR is connected to. Area 0 must be configured where multi-area OSPF exists, and for which all traffic sent between areas is generally required to traverse area 0, in order to ensure routing loops do not occur.

OSPF Network Advertisement




```
[RTA]ospf 1 router-id 1.1.1.1
[RTA-ospf-1]area 0
[RTA-ospf-1-area-0.0.0.0]network 192.168.1.0 0.0.0.255
```

- The network command defines the network to be advertised
- Route advertisements are forwarded based on areas.

Establishing of OSPF within an AS domain requires that each router that is to participate in OSPF first enable the OSPF process. This is achieved using the *ospf* [*process id*] command, where the process ID can be assigned and represents the process with which the router is associated. If routers are assigned different process ID numbers, separate link state databases will be created based on each individual process ID. Where no process ID is assigned, the default process ID of 1 will be used. The router ID can also be assigned using the command *ospf* [*process id*] [*router-id* <*router-id*>], where <router-id> refers to the ID that is to be assigned to the router, bearing in mind that a higher ID value represents the DR in broadcast and NBMA networks.

The parenthesis information reflects the ospf process and level at which ospf parameters can be configured, including the area to which each link (or interface) is associated. Networks that are to be advertised into a given area are determined through the use of the network command. The mask is represented as a wildcard mask for which a bit value of 0 represents the bits that are fixed (e.g. network id) and where the bit values in the mask represent a value of 1, the address can represent any value.

Configuration Validation

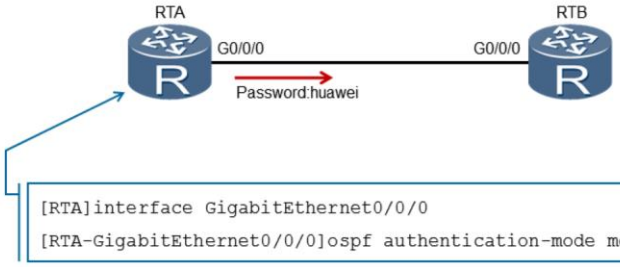
```
[RTA]display ospf peer

         OSPF Process 1 with Router ID 1.1.1.1
                 Neighbors

Area 0.0.0.0 interface 192.168.1.1(GigabitEthernet0/0/0)'s neighbors
Router ID: 2.2.2.2          Address: 192.168.1.2
   State: Full  Mode:Nbr is  Master  Priority: 1
   DR: 192.168.1.2  BDR: 192.168.1.1  MTU: 0
   Dead timer due in 40  sec
   Retrans timer interval: 5
   Neighbor is up for 00:00:31
   Authentication Sequence: [ 0 ]
```


Configuration of the neighbor relationship between OSPF peers is verified through the *display ospf peer* command. The attributes associated with the peer connection are listed to provide a clear explanation of the configuration. Important attributes include the area in which the peer association is established, the state of the peer establishment, the master/slave association for adjacency negotiation in order to reach the full state, and also the DR and BDR assignments which highlights that the link is associated with a broadcast network type.

OSPF Authentication



- OSPF supports two forms of authentication, simple password or cryptographic authentication.

HUAWEI

OSPF is capable of supporting authentication to ensure that routes are protected from malicious actions that may result from manipulation or damage to the existing OSPF topology and routes. OSPF allows for the use of simple authentication as well as cryptographic authentication, which provides enhanced protection against potential attacks.

Authentication is assigned on a per interface basis with the command for simple authentication of *ospf authentication-mode* { *simple* [[*plain*] <*plain-text*> | *cipher* <*cipher-text* >] | *null* } where plain applies a clear-text password, cipher a cipher-text password to hide the original contents, and null to indicate a null authentication.

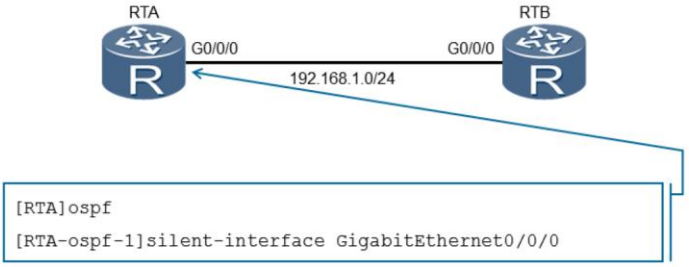
Cryptographic authentication is applied using the *ospf authentication-mode* { *md5* | *hmac-md5* } [*key-id* { *plain* <*plain-text* >| [*cipher*] <*cipher-text* >}] command. MD5 represents a cryptographic algorithm for securing authentication over the link, with its configuration demonstrated within the given example. The key identifies a unique authentication key ID of the cipher authentication of the interface. The key ID must be consistent with that of the peer.

Configuration Validation

```
<RTA>terminal debugging
<RTA>debugging ospf packet
Aug 19 2013 08:10:06.850.2+00:00 RTA RM/6/RMDEBUG:  Source Address:
192.168.1.1
Aug 19 2013 08:10:06.850.3+00:00 RTA RM/6/RMDEBUG:  Destination
Address: 224.0.0.5
……
Aug 19 2013 08:10:06.850.6+00:00 RTA RM/6/RMDEBUG:  Area: 0.0.0.0,
Chksum: 0
Aug 19 2013 08:10:06.850.7+00:00 RTA RM/6/RMDEBUG:  AuType: 02
Aug 19 2013 08:10:06.850.8+00:00 RTA RM/6/RMDEBUG:  Key(ascii): * *
* * * * * *
```


Where authentication is applied, it is possible to implement debugging on the terminal to view the authentication process. Since the debugging may involve many events, the *debugging ospf packet* command should be used to specify that debugging should only be performed for OSPF specific packets. As a result the authentication process can be viewed to validate that the authentication configuration has been successfully implemented.

OSPF Silent Interface



- The *silent-interface* command prevents an interface from forming neighbor relationships with peers.

It is often necessary to control the flow of routing information and limit the range for which such routing protocols can extend. This is particularly the case where connecting with external networks from whom knowledge of internal routes is to be protected. In order to achieved this, the silent interface command can be applied as a means to restrict all OSPF communication via the interface on which the command is implemented.

After an OSPF interface is set to be in the silent state, the interface can still advertise its direct routes. Hello packets on the interface, however, will be blocked and no neighbor relationship can be established on the interface. The command *silent-interface* [*interface-type interface-number*] can be used to define a specific interface that is to restrict OSPF operation, or alternatively the command *silent-interface all* can be used to ensure that all interfaces under a specific process be restricted from participating in OSPF.

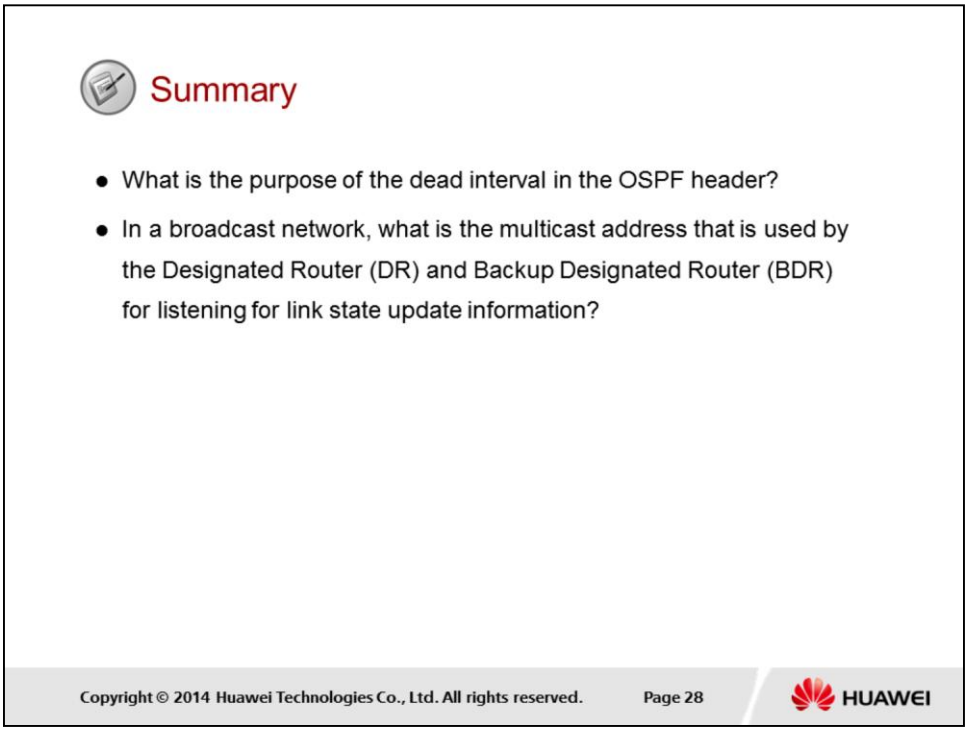
Configuration Validation

```
[RTA]display ospf 1 interface GigabitEthernet0/0/0

         OSPF Process 1 with Router ID 1.1.1.1
                 Interfaces

 Interface: 192.168.1.1 (GigabitEthernet0/0/0)
 Cost: 1       State: DR        Type: Broadcast    MTU: 1500
 Priority: 1
 Designated Router: 192.168.1.1
 Backup Designated Router: 0.0.0.0
 Timers: Hello 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit
Delay 1
 Silent interface, No hellos
```


The implementation of the silent interface on a per interface basis means that the specific interface should be observed to validate the successful application of the silent interface command. Through the *display ospf <process_id> interface <interface>* command, where the interface represents the interface to which the silent interface command has been applied, it is possible to validate the implementation of the silent interface.



1. The dead interval is a timer value that is used to determine whether the propagation of OSPF Hello packets has ceased. This value is equivalent to four times the Hello interval, or 40 seconds by default on broadcast networks. In the event that the dead interval counts down to zero, the OSPF neighbor relationship will terminate.
2. The DR and BDR use the multicast address 224.0.0.6 to listen for link state updates when the OSPF network type is defined as broadcast.

Thank you

www.huawei.com

DHCP Protocol Principles

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

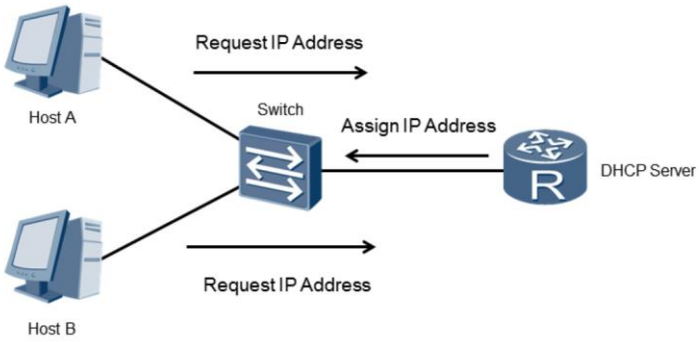
An enterprise network may often consist of a substantial number of host devices, each requiring network parameters in the form of IP addressing and additional network configuration information. Manual allocation is often a tedious and inaccurate business which can lead to many end stations facing address duplication or failure to reach services necessary for smooth network operation. DHCP is an application layer protocol that is designed to automate the process of providing such configuration information to clients within a TCP/IP network. DHCP therefore aids in ensuring correct addressing is allocated, and reduces the burden on administration for all enterprise networks. This section introduces the application of DHCP within the enterprise network.

Objectives

Upon completion of this section, trainees will be able to:

- Describe the function of DHCP in the enterprise network.
- Explain the leasing process of DHCP.
- Configure DHCP pools for address leasing.

DHCP Application In The Enterprise Network



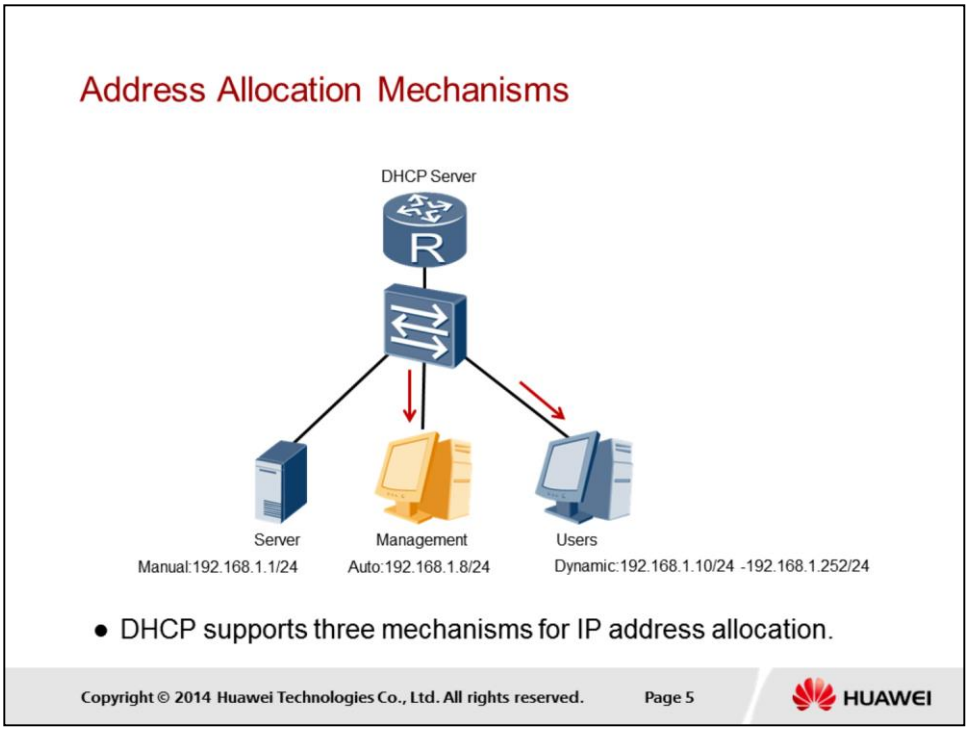
- Networks comprising of a large number of users requires a central management system for IP address allocation.

HUAWEI

Enterprise networks are often comprised of multiple end systems that require IP address assignment in order to connect with the network segment to which the end system is attached. For small networks, a minimal number of end systems attached to the network allows for simple management of the addressing for all end systems.

For medium and large-scale networks however, it becomes increasingly difficult to manually configure IP addresses with increased probability of duplication of addressing, as well as misconfiguration due to human error, and therefore the necessity to implement a centralized management solution over the entire network becomes ever more prominent. The Dynamic Host Configuration Protocol (DHCP) is implemented as a management solution to allow dynamic allocation of addresses for existing fixed and temporary end systems accessing the network domain.

In cases it is also possible that there may be more hosts than available IP addresses on a network. Some hosts cannot be allocated a fixed IP address and need to dynamically obtain IP addresses using the DHCP server. Only a few hosts on a network require fixed IP addresses.



DHCP supports three mechanisms for IP address allocation. The method of automatic allocation involves DHCP assigning a permanent IP address to a client. The use of dynamic allocation employs DHCP to assign an IP address to a client for a limited period of time or at least until the client explicitly relinquishes the IP address.

The third mechanism is referred to as manual allocation, for which a client's IP address is assigned by the network administrator, and DHCP is used only to handle the assignment of the manually defined address to the client. Dynamic allocation is the only one of the three mechanisms that allows automatic reuse of an address that is no longer needed by the client to which it was assigned. Thus, dynamic allocation is particularly useful for assigning an address to a client that will be connected to the network only temporarily, or for sharing a limited pool of IP addresses among a group of clients that do not need permanent IP addresses.

Dynamic allocation may also be a good choice for assigning an IP address to a new client being permanently connected to a network, where IP addresses are sufficiently scarce that addresses are able to be reclaimed when old clients are retired. Manual allocation allows DHCP to be used to eliminate the error-prone process of manually configuring hosts with IP addresses in environments where it may be more desirable to meticulously manage IP address assignment.

DHCP Messages

Message Types	Function
DHCP DISCOVER	Client broadcast used to locate available DHCP servers.
DHCP OFFER	Server responds to DHCPDISCOVER with an offer of configuration parameters
DHCP REQUEST	Client message to servers, either (a) requesting offered parameters from one server and implicitly declining offers from all others, (b) confirming the correctness of previously allocated address after, e.g., system reboot, or (c) extending the lease on a particular network address.
DHCP ACK	Server confirmation sent to the client with configuration parameters, including committed network address.
DHCP NAK	Server indicates to the client that client's requested network address cannot be assigned.
DHCP RELEASE	Client relinquishes the network address to the server and cancels the remaining lease.

A DHCP server and a DHCP client communicate with each other by exchanging a range of message types. Initial communication relies on the transmission of a DHCP Discover message. This is broadcast by a DHCP client to locate a DHCP server when the client attempts to connect to a network for the first time. A DHCP Offer message is then sent by a DHCP server to respond to a DHCP Discover message and carries configuration information.

A DHCP Request message is sent after a DHCP client is initialized, in which it broadcasts a DHCP Request message to respond to the DHCP Offer message sent by a DHCP server. A request message is also sent after a DHCP client is restarted, at which time it broadcasts a DHCP Request message to confirm the configuration, such as the assigned IP address. A DHCP Request message is also sent after a DHCP client obtains an IP address, in order to extend the IP address lease.

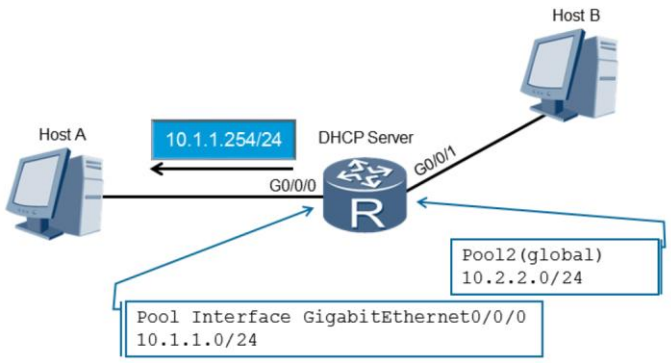
A DHCP ACK message is sent by a DHCP server to acknowledge the DHCP Request message from a DHCP client. After receiving a DHCP ACK message, the DHCP client obtains the configuration parameters, including the IP address. Not all cases however will result in the IP address being assigned to a client. The DHCP NAK message is sent by a DHCP server to in order reject the DHCP Request message from a DHCP client when the IP address assigned to the DHCP client expires, or in the case that the DHCP client moves to another network.

A DHCP Decline message is sent by a DHCP client, to notify the DHCP server that the assigned IP address conflicts with another IP address. The DHCP client will then apply to the DHCP server for another IP address.

A DHCP Release message is sent by a DHCP client to release its IP address. After receiving a DHCP Release message, the DHCP server assigns this IP address to another DHCP client.

A final message type is the DHCP Inform message, and is sent by a DHCP client to obtain other network configuration information such as the gateway address and DNS server address after the DHCP client has obtained an IP address.

Address Pools



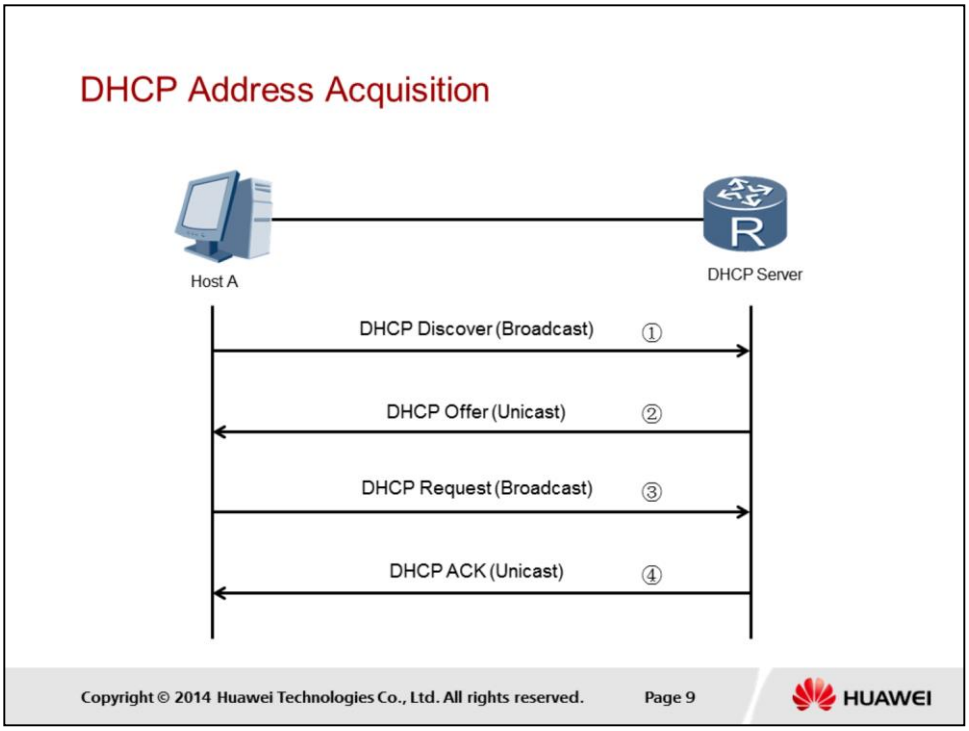
- Address pools can be either global or interface based.

  HUAWEI

The AR2200 and S5700 series devices can both operate as a DHCP server to assign IP addresses to online users. Address pools are used in order to define the addresses that should be allocated to end systems. There are two general forms of address pools which can be used to allocate addresses, the global address pool and the interface address pool.

The use of an interface address pool enables only end systems connected to the same network segment as the interface to be allocated IP addresses from this pool. The global address pool once configured allows all end systems associated with the server to obtain IP addresses from this address pool, and is implemented using the *dhcp select global* command to identify the global address pool. In the case of the interface address pool, the *dhcp select interface* command identifies the interface and network segment to which the interface address pool is associated.

The interface address pool takes precedence over the global address pool. If an address pool is configured on an interface, the clients connected to the interface obtain IP addresses from the interface address pool even if a global address pool is configured. On the S5700 switch, only logical VLANIF interfaces can be configured with interface address pools.



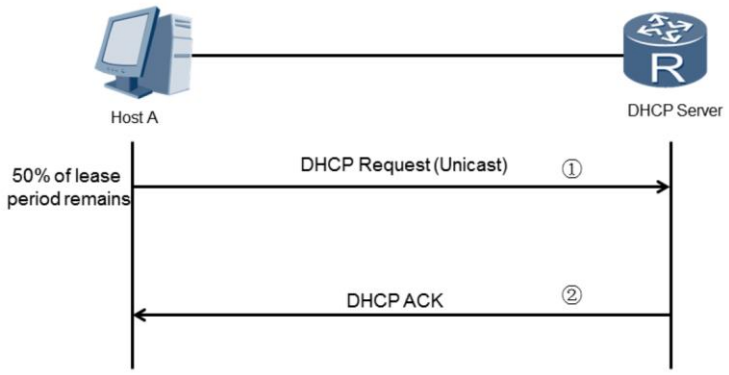
The acquisition of an IP address and other configuration information requires that the client make contact with a DHCP server and retrieve through request the addressing information to become part of the IP domain. This process begins with the IP discovery process in which the DHCP client searches for a DHCP server. The DHCP client broadcasts a DHCP Discover message and DHCP servers respond to the Discover message.

The discovery of one or multiple DHCP servers results in each DHCP server offering an IP address to the DHCP client. After receiving the DHCP Discover message, each DHCP server selects an unassigned IP address from the IP address pool, and sends a DHCP Offer message with the assigned IP address and other configuration information to the client.

If multiple DHCP servers send DHCP Offer messages to the client, the client accepts the first DHCP Offer message received. The client then broadcasts a DHCP Request message with the selected IP address. After receiving the DHCP Request message, the DHCP server that offers the IP address sends a DHCP ACK message to the DHCP client. The DHCP ACK message contains the offered IP address and other configuration information.

Upon receiving the DHCP ACK message, the DHCP client broadcasts gratuitous ARP packets to detect whether any host is using the IP address allocated by the DHCP sever. If no response is received within a specified time, the DHCP client uses this IP address. If a host is using this IP address, the DHCP client sends the DHCP Decline packet to the DHCP server, reporting that the IP address cannot be used, following which the DHCP client applies for another IP address.

DHCP Lease Renewal



- DHCP initiates an IP lease renewal process when a lease period of 50% remains.

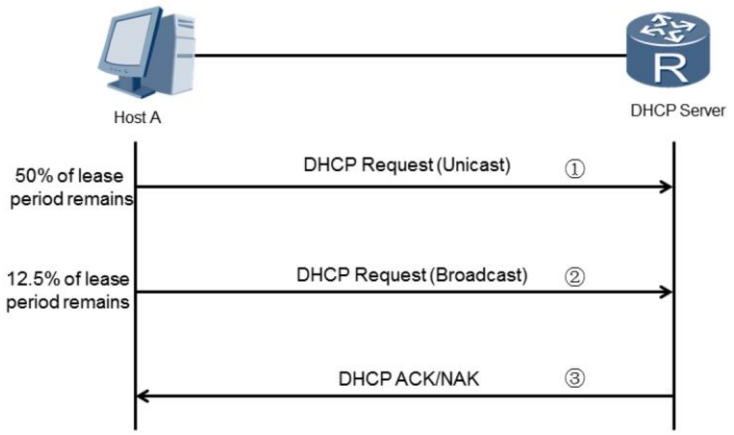
HUAWEI

After obtaining an IP address, the DHCP client enters the binding state. Three timers are set on the DHCP client to control lease update, lease rebinding, and lease expiration. When assigning an IP address to a DHCP client, a DHCP server specifies values for the timers.

If the DHCP server does not set the values for the timers, the DHCP client uses the default values. The default values define that when 50% of the lease period remains, the release renewal process should begin, for which a DHCP client is expected to renew its IP address lease. The DHCP client automatically sends a DHCP Request message to the DHCP server that has allocated an IP address to the DHCP client.

If the IP address is valid, the DHCP server replies with a DHCP ACK message to entitle the DHCP client a new lease, and then the client re-enters the binding state. If the DHCP client receives a DHCP NAK message from the DHCP server, it enters the initializing state.

DHCP Rebinding Expiry



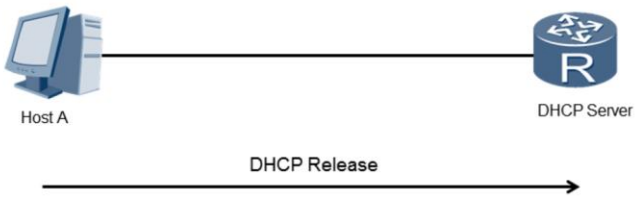
- Rebinding will occur if the lease is not renewed in time.

HUAWEI

After the DHCP client sends a DHCP Request message to extend the lease, the DHCP client remains in an updating state and waits for a response. If the DHCP client does not receive a DHCP Reply message from the DHCP server after the DHCP server rebinding timer expires which by default occurs when 12.5% of the lease period remains, the DHCP client assumes that the original DHCP server is unavailable and starts to broadcast a DHCP Request message, for which any DHCP server on the network can reply with a DHCP ACK or NAK message.

If the received message is a DHCP ACK message, the DHCP client returns to the binding state and resets the lease renewal timer and server binding timer. If all of the received messages are DHCP NAK messages, the DHCP client goes back to the initializing state. At this time, the DHCP client must stop using this IP address immediately and request a new IP address.

IP Address Release

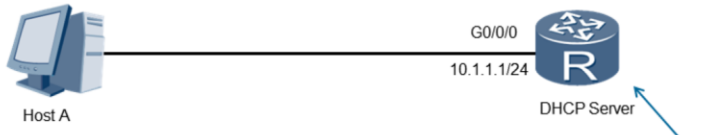


- DHCP will result in the release of an IP address if the client fails to renew the IP address before the lease expiry.

HUAWEI

The lease timer is the final timer in the expiration process, and if the DHCP client does not receive a response before the lease expiration timer expires, the DHCP client must stop using the current IP address immediately and return to the initializing state. The DHCP client then sends a DHCP DISCOVER message to apply for a new IP address, thus restarting the DHCP cycle.

DHCP Interface Pool Configuration




```
[Huawei]dhcp enable
[Huawei]interface GigabitEthernet0/0/0
[Huawei-GigabitEthernet0/0/0]dhcp select interface
[Huawei-GigabitEthernet0/0/0]dhcp server dns-list 10.1.1.2
[Huawei-GigabitEthernet0/0/0]dhcp server excluded-ip-address
10.1.1.2
[Huawei-GigabitEthernet0/0/0]dhcp server lease day 3
```


HUAWEI

There are two forms of pool configuration that are supported in DHCP, these include defining a global pool or an interface based pool. The *dhcp select interface* command is used to associate an interface with the interface address pool in order to provide configuration information to connected hosts. The example demonstrates how interface Gigabit Ethernet 0/0/0 has been assigned as part of an interface address pool.

DHCP Configuration Validation

```
[Huawei]display ip pool interface GigabitEthernet0/0/0
  Pool-name      : GigabitEthernet0/0/0
  Pool-No        : 0
  Lease          : 3 Days 0 Hours 0 Minutes
  Domain-name    : huawei.com
  DNS-Server0    : 10.1.1.2
  NBNS-Server0   : -
  Netbios-type   : -
  Position       : Interface       Status      : Unlocked
  Gateway-0      : 10.1.1.1
  Mask           : 255.255.255.0
  VPN instance   : --
 ------------------------------------------------------------------
  Start       End     Total Used  Idle(Expired) Conflict Disable
 ------------------------------------------------------------------
 10.1.1.1  10.1.1.254  253   1        251(0)         0       1
```


Each DHCP server will define one or multiple pools which may be associated globally or with a given interface. For determining the pool attributes associated with an interface, the *display ip pool interface <interface>* command is used. The DHCP pool will contain information including the lease period for each IP address that is leased, as well as the pool range that is supported. In the event that other attributes are supported for DHCP related propagation to clients such as with the IP gateway, subnet mask, and DNS server, these will also be displayed.

DHCP Global Pool Configuration


```
[Huawei]dhcp enable
[Huawei]ip pool pool2
Info: It's successful to create an IP address pool.
[Huawei-ip-pool-pool2]network 10.2.2.0 mask 24
[Huawei-ip-pool-pool2]gateway-list 10.2.2.1
[Huawei-ip-pool-pool2]lease day 1
[Huawei-ip-pool-pool2]quit
[Huawei]interface GigabitEthernet0/0/1
[Huawei-GigabitEthernet0/0/1]dhcp select global
```

- Establishment of an address pool and associated parameters is implemented on the DHCP server.

HUAWEI

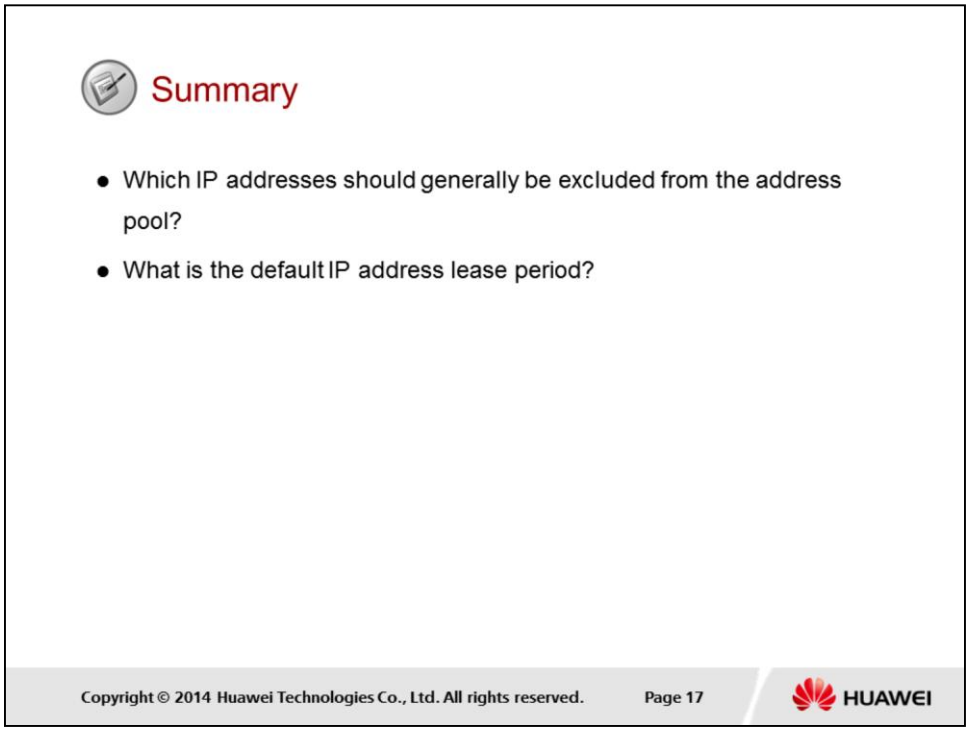
The example demonstrates the DHCP configuration for a global address pool that is assigned to the network 10.2.2.0. The *dhcp enable* command is the prerequisite for configuring DHCP-related functions, and takes effect only after the *dhcp enable* command is run. A DHCP server requires the *ip pool* command be configured in the system view to create an IP address pool and set IP address pool parameters, including a gateway address, the IP address lease period etc. The configured DHCP server can then assign IP addresses in the IP address pool to clients.

A DHCP server and its client may reside on different network segments. To enable the client to communicate with the DHCP server, the *gateway-list* command is used to specify an egress gateway address for the global address pool of the DHCP server. The DHCP server can then assign both an IP address and the specified egress gateway address to the client. The address is configured in dotted decimal notation for which a maximum of eight gateway addresses, separated by spaces, can be configured.

DHCP Configuration Validation

```
[Huawei]display ip pool
 ------------------------------------------------------------
  Pool-name      : pool2
  Pool-No        : 0
  Position       : Local           Status           : Unlocked
  Gateway-0      : 10.2.2.1
  Mask           : 255.255.255.0
  VPN instance   : --
  IP address Statistic
    Total        :253
    Used         :1          Idle         :252
    Expired      :0          Conflict     :0          Disable   :0
```


The information regarding a pool can be also observed through the used of the *display ip pool* command. This command will provide an overview of the general configuration parameters supported by a configured pool, including the gateway and subnet mask for the pool, as well general statistics that allow an administrator to monitor the current pool usage, to determine the number of addresses allocated, along with other usage statistics.



1. IP addresses that are used for server allocation such as any local DNS servers in order to avoid address conflicts.
2. The default lease period for DHCP assigned IP addresses is set at a period equal to one day.

Thank you

www.huawei.com

FTP Protocol Principles

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

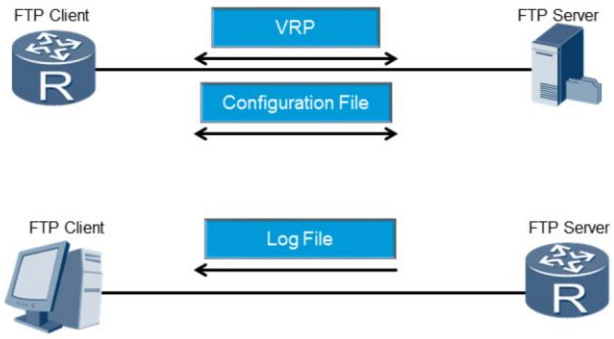
Early development of standards introduced the foundations of a file transfer protocol, with the aim of promoting the sharing of files between remote locations that were not impacted by variations in file storage systems among hosts. The resulting FTP application was eventually adopted as part of the TCP/IP protocol suite. The FTP service remains an integral part of networking as an application for ensuring the reliable and efficient transfer of data, commonly implemented for effective backup and retrieval of files and logs, thereby improving overall management of the enterprise network. This section therefore introduces the means for engineers and administrators to implement FTP services within Huawei products.

Objectives

Upon completion of this section, trainees will be able to:

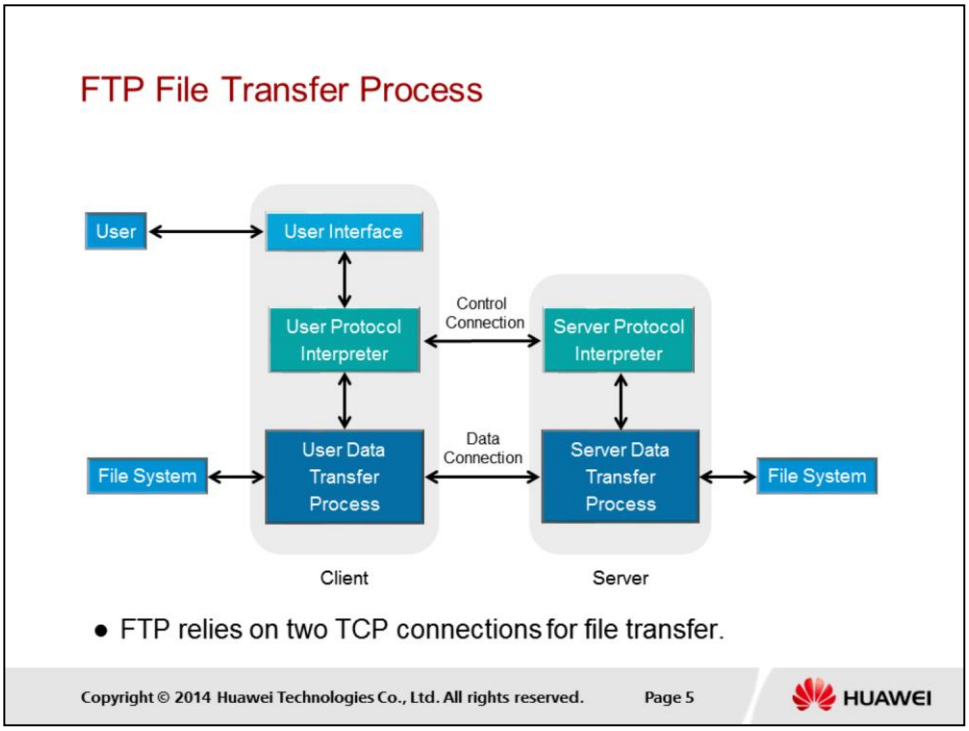
- Explain the file transfer process of FTP.
- Configure the FTP service on supporting Huawei devices.

FTP Application In The Enterprise Network



- FTP provides an effective means for backup and retrieval of important files.

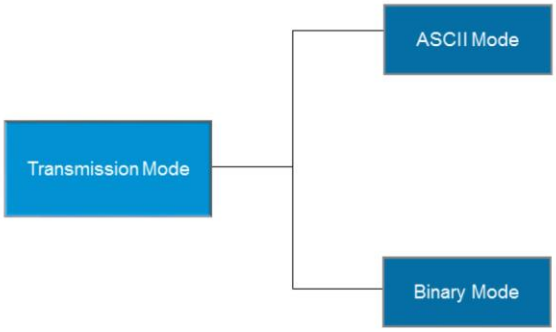
The implementation of an FTP server within the enterprise network allows for effective backup and retrieval of important system and user files, which may be used to maintain the daily operation of an enterprise network. Typical examples of how an FTP server may be applied include the backing up and retrieval of VRP image and configuration files. This may also include the retrieval of log files from the FTP server to monitor the FTP activity that has occurred.



The transfer of FTP files relies on two TCP connections. The first of these is a control connection which is set up between the FTP client and the FTP server. The server enables common port 21 and then waits for a connection request from the client. The client then sends a request for setting up a connection to the server. A control connection always waits for communication between the client and the server, transmits related commands from the client to the server, as well as responses from the server to the client.

The server uses TCP port 20 for data connections. Generally, the server can either open or close a data connection actively. For files sent from the client to the server in the form of streams, however, only the client can close a data connection. FTP transfers each file in streams, using an End of File (EOF) indicator to identify the end of a file. A new data connection is therefore required for each file or directory list to be transferred. When a file is being transferred between the client and the server, it indicates that a data connection is set up.

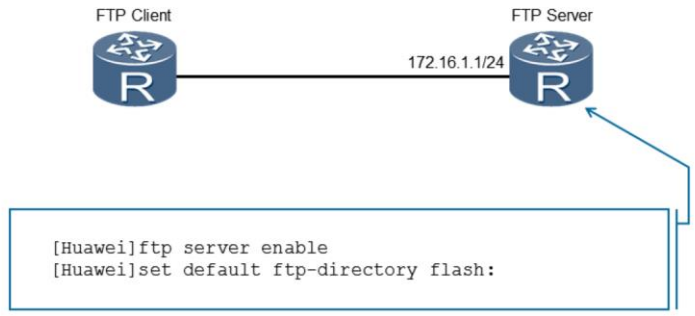
FTP Basic Concepts



- Transmission modes define the format of data before it is carried between the sender and receiver.

There are two FTP transmission modes which are supported by Huawei, these are ASCII mode and binary mode. ASCII mode is used for text, in which data is converted from the sender's character representation to "8-bit ASCII" before transmission. Put simply, ASCII characters are used to separate carriage returns from line feeds. In binary mode the sender sends each file byte for byte. This mode is often used to transfer image files and program files for which characters can be transferred without format converting.

FTP Service Establishment

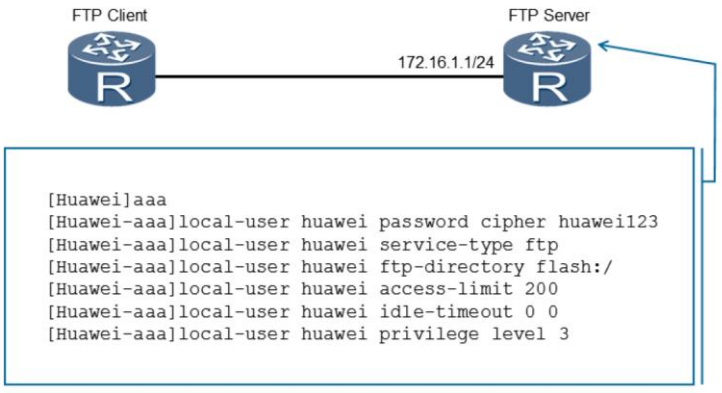


- The FTP service must be enabled and a default FTP directory specified for file handling.

HUAWEI

Implementing the FTP service is achievable on both the AR2200 series router and S5700 series switch, for which the service can be enabled through the *ftp server enable* command. After the FTP server function is enabled, users can manage files in FTP mode. The *set default ftp-directory* command sets the default working directory for FTP users. Where no default FTP working directory is set, the user will be unable to log into the router, and will instead be prompted with a message notifying that the user has no authority to access any working directory.

FTP Service User Establishment



- User accounts can be implemented to identify users and as well as apply specific permissions for each user individually.

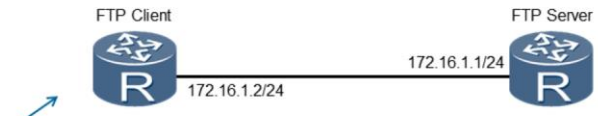
HUAWEI

Access to the FTP service can be achieved by assigning individual user login to manage access on a per user basis. AAA is used to configure local authentication and authorization. Once the AAA view is entered, the local user can be created, by defining a user account and password. The account is capable of associating with a variety of services, which are specified using the *service-type* command, to allow for the ftp service type to be supported by AAA.

If the ftp directory of the user should vary from the default directory, the *ftp-directory* command can be used to specify the directory for the user. If the number of active connections possible with a local user account is to be limited, the *access-limit* command can be applied. This can range from 1 to 800, or unlimited where an access limit is not applied.

The configuration of an idle timeout helps to prevent unauthorized access in the event that a session window is left idle for a period of time by a user. The *idle timeout* command is defined in terms of minutes and seconds, with an idle timeout of 0 0 representing that no timeout period is applied. Finally the privilege level defines the authorized level of the user in terms of the commands that can be applied during ftp session establishment. This can be set for any level from 0 through to 15, with a greater value indicating a higher level of the user.

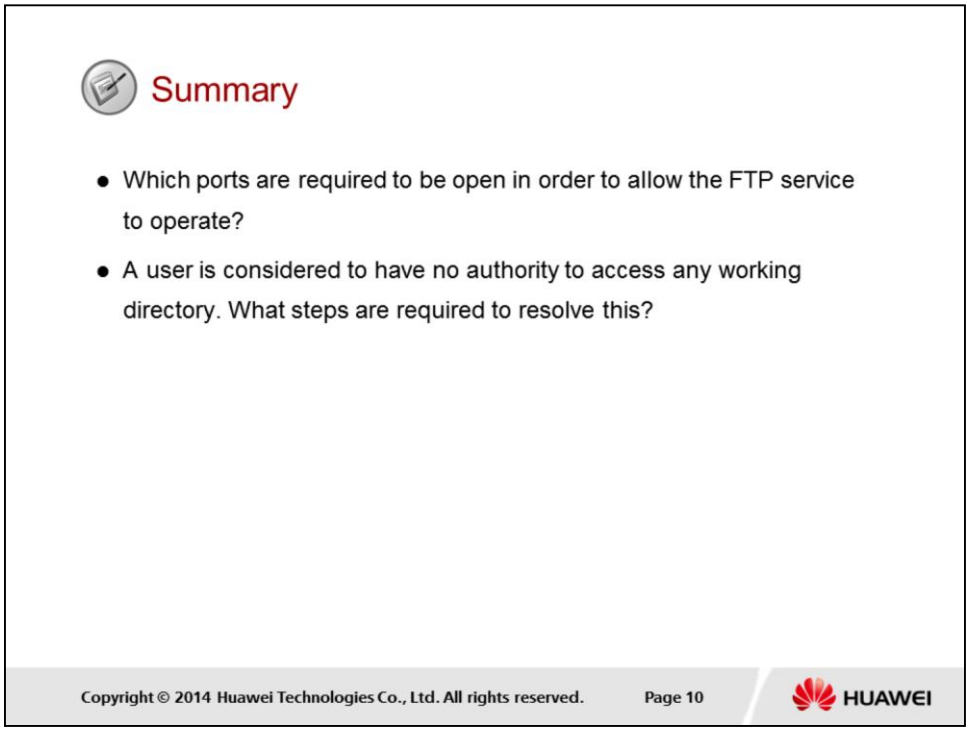
FTP User Configuration




```
<Huawei>ftp 172.16.1.1
Trying 172.16.1.1 ...
Press CTRL+K to abort
Connected to 172.16.1.1.
220 FTP service ready.
User(172.16.1.1:(none)):huawei
331 Password required for huawei.
Enter password:
230 User logged in.
[ftp]binary
200 Type set to I.
[ftp]get vrp.cc
```


HUAWEI

Following configuration of the FTP service on the FTP server, it is possible for users to establish a connection between the client and the server. Using the *ftp* command on the client will establish a session through which the AAA authentication will be used to validate the user using AAA password authentication. If correctly authenticated, the client will be able to configure as well as send/retrieve files to and from the FTP server.



1. In order for the control connection and data connection of the FTP service to be established successfully, TCP ports 20 and 21 must be enabled.
2. When a user is considered to have no authority to access any working directory, a default FTP directory needs to be defined. This is done by using the command *set default ftp-directory <directory location>*, where the directory name may be, for example, the system flash.

Thank you

www.huawei.com

Telnet Protocol Principles

HUAWEI TECHNOLOGIES CO., LTD.

Foreword

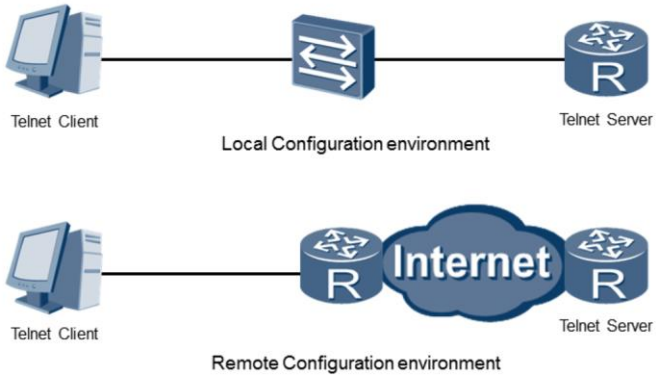
As the enterprise network expands, supported devices may exist over a larger geographical distance due to the presence of branch offices that are considered part of the enterprise domain, and that require remote administration. Additionally the administration of the network is often managed from a central management location from which all devices are monitored and administered. In order to facilitate this administration, the telnet protocol, one of the earliest protocols to be developed, is applied to managed devices. The principles surrounding the protocol and it's implementation are introduced in this section.

Objectives

Upon completion of this section, trainees will be able to:

- Explain the application and principles surrounding telnet
- Establish the telnet service on supporting Huawei devices.

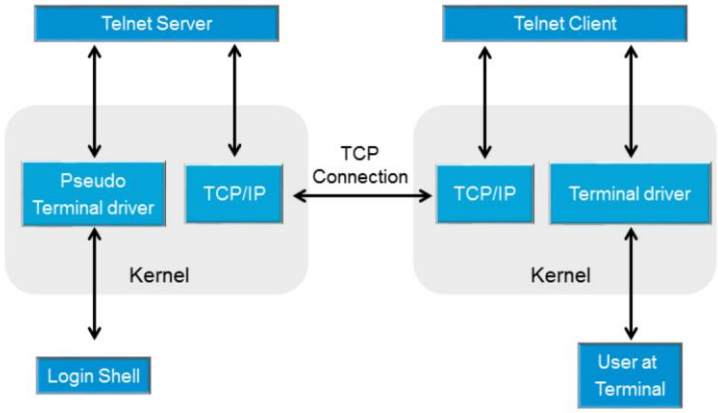
Telnet Application



- Telnet represents a bidirectional text based terminal emulation program for use over local and remote networks.

The Telecommunication Network Protocol (Telnet) enables a terminal to log in remotely to any device which is capable of operating as a telnet server, and provides an interactive operational interface via which the user can perform operations, in the same manner as is achieved locally via a console connection. Remote hosts need not be connected directly to a hardware terminal, allowing instead for users to take advantage of the ubiquitous capabilities of IP in order to remotely manage devices from almost any location in the world.

Telnet Client/Server Model



- The Telnet architecture demonstrates how user keystrokes are interpreted by terminal drivers before delivery over TCP ensues.

HUAWEI

Telnet operates on a client/server model principle for which a telnet TCP connection is established between a user port and the server telnet port, which by default is assigned as port 23. The server listens on this well known port for such connections. A TCP connection is full duplex and identified by the source and destination ports. The server can engage in many simultaneous connections involving its well known port and user ports that are assigned from a non well-known port range.

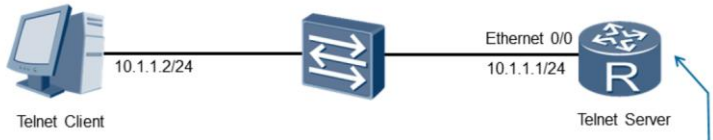
The telnet terminal drivers interpret the keystrokes of users and translates these to a universal character standard, based on a network virtual terminal (NVT) which operates as a form of virtual intermediary between systems, following which the transmission via the TCP/IP connection to the server is performed. The server decodes the NVT characters and passes the decoded characters to a pseudo terminal driver which exists to allow the operating system to receive the decoded characters.

Authentication Mode

Authentication Mode	Description
None	Login without authentication
AAA	AAA authentication
Password	Authentication through the password of a user terminal interface

Access to the telnet service commonly involves authentication of the user before access is granted. There are three main modes that are defined for telnet authentication.

Telnet Configuration



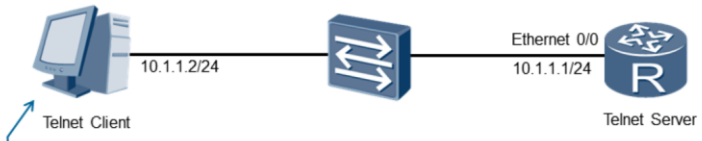

```
[Huawei]interface Ethernet 0/0/0
[Huawei-Ethernet0/0/0]ip address 10.1.1.1 24
[Huawei]user-interface vty 0 4
[Huawei-ui-vty0-4]authentication-mode password
[Huawei-ui-vty0-4]set authentication password cipher
Enter Password(<8-128>): huawei12
```

- Telnet requires authentication be applied to the virtual teletype interface before a connection can be established.

HUAWEI

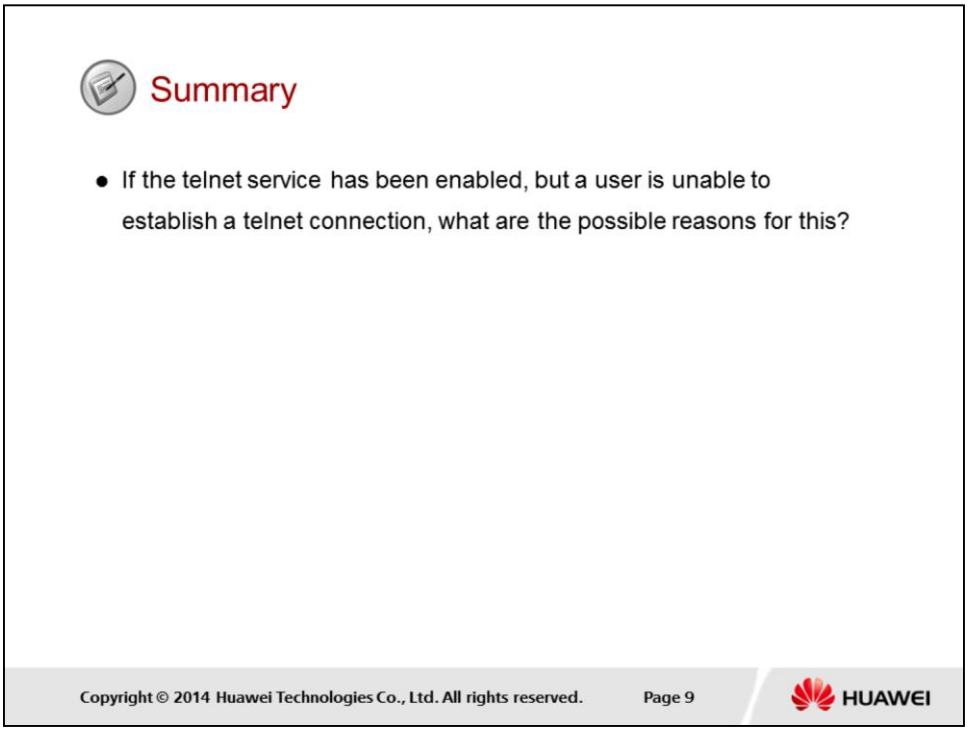
Establishment to a device operating as a telnet server commonly uses a general password authentication scheme which is used for all users connecting to the user vty interface. Once IP connectivity is established through the implementation of a suitable addressing scheme, the *authentication-mode password* command set is implemented for the vty range, along with the password to be used.

Telnet Configuration




```
<Host>telnet 10.1.1.1
Trying 10.1.1.1 ...
Press CTRL+K to abort
Connected to 10.1.1.1 ...
Login authentication
Password:
Info: The max number of VTY users is 5, and the number
      of current VTY users on line is 1.
      The current login time is 2013-04-19 16:32:00.
<Huawei>
```


Following configuration of the remote device that is to operate as a telnet server, the client is able to establish a telnet connection through the *telnet* command, and receive the prompt for authentication. The authentication password should match the password implemented on the telnet server as part of the prior password authentication configuration. The user will be then able to establish a remote connection via telnet to the remote device operating as a telnet server and emulate the command interface on the local telnet client.



1. If a user is unable to establish a telnet connection, the user should verify the device supporting the telnet service is reachable. If the device can be reached, the password should be verified. If the password is considered to be correct, the number of users currently accessing the device via telnet should be checked. If it is necessary to extend the number of users accessing the device through telnet, the user-interface maximum-vty <0-15> command should be used, where 0-15 denotes the number of supported users.

Thank you

www.huawei.com